Discrete Logic Modelling as a Means to Link Protein Signalling Networks with Functional Analysis of Mammalian Signal Transduction

Molecular Systems Biology 5; Article number 331; doi:10.1038/msb.2009.87
Citation: ***Molecular Systems Biology*** **5:331**

www.molecularsystemsbiology.com

molecular systems biology

Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction

Julio Saez-Rodriguez[1,2,3,5], Leonidas G Alexopoulos[1,2,3,5,6], Jonathan Epperlein[1,2], Regina Samaga[4], Douglas A Lauffenburger[1,3], Steffen Klamt[4] and Peter K Sorger[1,2,3,*]

[1] Center for Cell Decision Processes, Boston, MA, USA, [2] Department of Systems Biology, Harvard Medical School, Boston, MA, USA, [3] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA and [4] Department of Systems Biology, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany
[5] These authors contributed equally to this work
[6] Present address: Department of Mechanical Engineering, National Technical University of Athens, Zografou 15780, Greece
* Corresponding author. Department of Systems Biology, Harvard Medical School, Warren Alpert 438, 200 Longwood Avenue, Boston, MA 02115, USA. Tel.: +1 617 432 6901/Ext. 6902; Fax: +1 617 432 5012; E-mail: sbpipeline@hms.harvard.edu

Received 19.3.09; accepted 28.10.09

Large-scale protein signalling networks are useful for exploring complex biochemical pathways but do not reveal how pathways respond to specific stimuli. Such specificity is critical for understanding disease and designing drugs. Here we describe a computational approach—implemented in the free CNO software—for turning signalling networks into logical models and calibrating the models against experimental data. When a literature-derived network of 82 proteins covering the immediate-early responses of human cells to seven cytokines was modelled, we found that training against experimental data dramatically increased predictive power, despite the crudeness of Boolean approximations, while significantly reducing the number of interactions. Thus, many interactions in literature-derived networks do not appear to be functional in the liver cells from which we collected our data. At the same time, CNO identified several new interactions that improved the match of model to data. Although missing from the starting network, these interactions have literature support. Our approach, therefore, represents a means to generate predictive, cell-type-specific models of mammalian signalling from generic protein signalling networks.
Molecular Systems Biology **5**: 331; published online 1 December 2009; doi:10.1038/msb.2009.87
Subject Categories: computational methods; signal transduction
Keywords: logical modelling; protein networks; signal transduction

Introduction

Successful identification of transmembrane receptors, intracellular signalling proteins, and transcription factors mediating the responses of cells to intra- and extracellular ligands has generated a wealth of information about the biochemistry of signal transduction (Hanahan and Weinberg, 2000). However, accumulation of molecular detail does not automatically yield improved understanding of the ways in which signalling circuits process complementary and opposing inputs to control diverse physiological responses. For this we require network-level perspectives. One approach to organizing data on large groups of genes and proteins is to create interaction networks using either of two related methods (Pieroni *et al*, 2008; Cusick *et al*, 2009). One infers connectivity directly from systematic two-hybrid, affinity purification/mass spectrometry and related high-throughput data (Rual *et al*, 2005), and the second culls interactions from the literature (Bauer-Mehren *et al*, 2009). Literature curation can be performed by expert readers or automatically using 'bibliome' mining software (Zhou and He, 2008). The resulting information is usually represented as a node–edge graph and stored in public databases such as Pathway Commons ((www.pathwaycommons.org); see Pathguide for a comprehensive list (Bader *et al*, 2006)) or in proprietary softwares from companies such as Ingenuity (Redwood City, CA, USA). Such node–edge graphs are often redrawn to create

graphically pleasing posters and Web-accessible pictograms (e.g., Biocarta).

As outlined by Pieroni *et al* (2008), protein node-edge graphs can be classified into two families: large-scale protein interaction networks (PINs—or 'interactomes'), which depict interactions between protein nodes (species) as undirected edges, and protein signalling networks (PSNs) whose edges have a sign (activating or inhibitory) and directionality (enzyme–substrate relationships). PINs are usually created using data from bibliome mining (Chatr-Aryamontri *et al*, 2007; Kerrien *et al*, 2007), large-scale affinity purification (Köcher and Superti-Furga, 2007), protein arrays (MacBeath and Schreiber, 2000), and two-hybrid screening (Rual *et al*, 2005) or genetic interactions (Jansen *et al*, 2003), whereas PSNs are most commonly assembled by expert annotation of the literature (Ma'ayan *et al*, 2005). However, PSNs can also be assembled using 'reverse engineering' methods such as Bayesian network analysis (Sachs *et al*, 2005) or inferred systems of differential equations (Nelander *et al*, 2008). The utility of PINs and PSNs is increased by incorporation of Gene Ontology (GO) tags ((Harris *et al*, 2004) and information from the KEGG database (Kanehisa *et al*, 2004). Nodes and edges can also be referenced to standardized ontologies such as BioPAX. The topologies of PINs and PSNs have been studied from an information-theoretic standpoint, with the goal of extracting principles of network design (Barabási and Oltvai, 2004; Pieroni *et al*, 2008). Moreover, overlay of expression data on PINs and PSNs makes it possible to explore differential activation of sub-networks in various conditions and cell types (Luscombe *et al*, 2004; Bossi and Lehner, 2009); annotating PINs with data has proven useful in predicting outcomes in breast cancer patients (Taylor *et al*, 2009).

Despite these developments, protein networks inferred purely from data and those assembled from the literature suffer from significant and complementary weaknesses: reverse-engineered networks ignore a wealth of existing mechanistic information about individual proteins and reaction intermediates, whereas literature-based networks are too disconnected from functional data to reveal input–output relationships. Thus, even the most comprehensive PINs and PSNs do not capture the logic of cellular biochemistry and—critically—cannot predict the responses of cells to specific biological stimuli. To determine whether a particular interaction network is consistent with a set of experimental data, we require a means to compute the state or output of a network given a set of input conditions. For example, it might be clear that two nodes in a signed directed graph have a positive effect on a downstream node, but a graph alone cannot specify whether the target is active in the presence of either node or only when both are present.

One means to convert a graph into a computable model is to encode it as a system of differential equations. This generates a detailed and biochemically realistic representation, but at the cost of many free parameters, which must be estimated. When the number of species in the network is large (tens to hundreds), parameter estimation becomes very challenging (Aldridge *et al*, 2006). An alternative approach is to depict the pathway as a logical model in which gates specify how outputs are related to inputs. Two-state discrete (Boolean) logic is the simplest logical representation and has no free parameters: logical models covering the same set of nodes differ only in topology. Boolean modelling has previously been applied to biological regulatory and signalling networks (Kauffman, 1969; Thomas and D'Ari, 1990; Huang and Ingber, 2000; Thomas and Kaufman, 2001; de Jong, 2002; Chaves *et al*, 2005; Fauré *et al*, 2006; Fisher and Henzinger, 2007; Gupta *et al*, 2007; Saez-Rodriguez *et al*, 2007; Zhang *et al*, 2008; Samaga *et al*, 2009), but a significant challenge in modelling biochemical pathways using Boolean logic has not yet been addressed: optimizing models against experimental data. In the absence of data-dependent optimization it is difficult to determine whether logical models can make accurate biological predictions and yield new insight.

In this paper we attempt to span the divide between interaction-focused networks (PINs and PSNs) and functional studies of cellular biochemistry, and between literature and data-centric approaches to network modelling. We describe a method for assembling Boolean logic models from a PSN and calibrating models (determining the optimal topology) against functional data using a newly developed and freely available software package (CellNetOptimizer; CNO). CNO first compresses PSNs to remove non-identifiable elements and then converts them into a hypergraph representing a superposition of all Boolean models compatible with the PSN. This superstructure of models is then trained against experimental data by minimizing an objective function that quantifies the difference between data and simulation while penalizing model size. Finally, optimized models are used to predict new results and are mined for biological insight. We illustrate CNO with a toy pathway and then apply it to an 85-protein signal transduction circuit that mediates immediate-early signalling downstream of seven cytokine and growth factor receptors in human liver cells. The training data for this PSN comprises a set of $\sim$1000 biochemical measurements that conformed to a cue-signal-response (CSR) paradigm (Gaudet *et al*, 2005). HepG2 hepatocellular carcinoma cells were exposed to combinations of extracellular ligands and small-molecule inhibitors, and the phosphorylation states or abundance of adaptor proteins, intracellular kinases, transcription factors, and so on were measured using high-throughput biochemical assays (Alexopoulos *et al*, in preparation). The training data were stored and processed using an updated version of our recently developed data management application, DataRail (Saez-Rodriguez *et al*, 2008). Lastly, a set of predictions were made using the trained model and the predictions confirmed experimentally (using data unique to this paper).

We show here that data-optimized Boolean models of cell signalling have considerably fewer connections than the literature-based PSNs from which they are derived, but have superior false-positive–false-negative rates, and do a better job of predicting data absent from the training set. Thus, the radical simplification involved in modelling cellular biochemistry using a discrete two-state Boolean formalism does not preclude optimization of model topology. Training served to eliminate many interactions present in the starting PSN and the eventual size of data-optimized Boolean models was remarkably robust to changes in the size penalty imposed during training. In HepG2 cells, interactions that were eliminated included canonical interactions between growth and inflammatory factors. Removal of these interactions did

not appear to be an artefact of our approach because the interactions were retained in models of other cell types. The fit of optimized models could be further improved by adding a limited number of links absent from the starting PSN. Newly added links did not appear to be spurious because they have support in the literature.

Results

Assembly of a Boolean model

We wrote the CNO software in MATLAB as a means to assemble and train Boolean models of biological pathways and then tested CNO on a plausible but imaginary 'toy' pathway and an associated set of synthetic data. The toy pathway comprised a subset of the intracellular signalling proteins known to be activated by epidermal growth factor or tumour necrosis factor (TNF) receptors in mammalian cells (EGFR and TNFR) and was represented as a signed directed graph (PSN) having a total of 18 nodes (Figure 1A). A Boolean model (the 'reference' model) was assembled manually consistent with the network graph and then used to compute the discretized activities of four signalling proteins following ligand stimulation in the presence or absence of a small-molecule inhibitor of PI3K (e.g., ZSTK474; Yaguchi *et al*, 2006) or Raf kinase (e.g., Sorafenib; Hotte and Hirte, 2002). The synthetic data corresponded to levels of phosphorylation at activating sites for AKT and ERK, nuclear translocation of NFκB, and cleavage of caspase-8 (Figure 1B). CNO was used to reconstruct a Boolean model from the PSN and synthetic data without knowledge of the reference model. The fidelity of the reconstruction was judged by the degree of similarity between the CNO-based reconstruction and the original reference model.

The first step in model assembly was compression of the pathway graph to remove non-identifiable elements. The nodes and edges subjected to experimental manipulation or measurement were labelled as 'designated', while the remaining nodes were labelled as 'undesignated'. Designated nodes in the toy model included TGFα and TNFα ligands, kinases that were subject to inhibition by small-molecule drugs, antibodies or RNAi, and signalling proteins whose levels, states, or activities were directly measured (Figure 1B). Compression of undesignated elements involved the application of three procedures. First, CNO automatically flagged for omission all species and interactions that did not alter any designated species. These lay on terminal branches of the pathway graph and corresponded to non-observables in systems theory (Kremling and Saez-Rodriguez, 2007). Species whose states were not affected by any of the inputs and perturbations (the ligands and inhibitors in this case) were also eliminated; these corresponded to non-controllable elements. Second, CNO compressed cascades in which a series of undesignated nodes and edges impinged on a designated node; these typically involved linear cascades or subnetworks of converging or diverging interactions in which no measurements or manipulations were made; the three situations in which this arises are illustrated in Figure 1C. Third, CNO retained undesignated nodes that remained after application of the preceding two procedures; this occurred when several links converged on a single undesignated element and then diverged from it (Figure 1C). Compression of such subnetworks can create internally inconsistent logic.

Compression of non-observable pathways (application of procedure one) is illustrated in the toy graph in Figure 1D by *GSK3*. The state of *GSK3* was not measured and its activity was not subjected to manipulation. CNO, therefore, removed both GSK3 and the AKT→GSK3 link. Application of the second procedure is illustrated by compression of the path TGFα→EGFR→Shc→Grb2/Sos→Ras→Raf into TGFα→Raf. The alternative path from TGFα to Raf via Shc (TGFα→EGFR→Grb2/Sos→Ras→Raf) was also compressed into TGFα→Raf, and thus the two parallel paths were automatically reduced to TGFα→Raf. If compression results in two parallel paths that share a starting and an ending node but have different sign, CNO keeps both. Overall, CNO compressed the toy graph of 18 nodes into a graph with eight designated nodes (Figure 1D). CNO keeps track of all nodes and edges eliminated during compression, making it possible to decompress the model following calibration. This serves to increase the intelligibility of the network because it re-casts the model in terms of known biochemical causality (e.g., Raf→MEK→ERK rather than Raf→ERK) and simplifies another round of modelling based on additional data and new designated species.

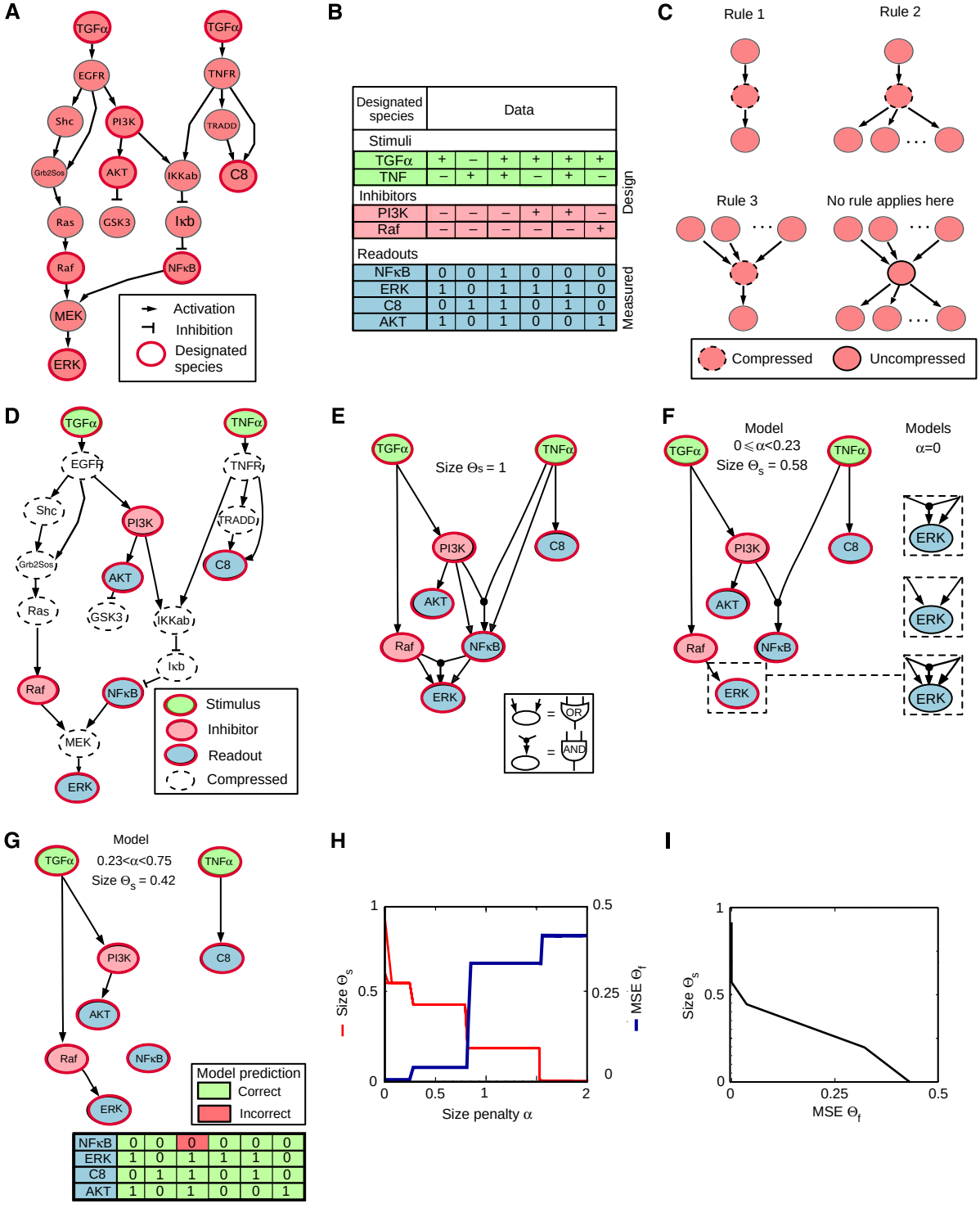
Next we created a superstructure of Boolean models having all possible logic gates compatible with the compressed graph. The superstructure was represented as a hypergraph using the sum-of-products formalism (see section Materials and methods and reference Klamt *et al*, 2006) in which multiple OR and AND gates are combined, and inhibition is encoded using NOT operators. For example, in the compressed toy graph Raf and NFκB are upstream of ERK (MEK was removed by compression), but the logic of their interaction is not known *a priori* and CNO, therefore, encodes the relationship by assuming that both upstream molecules are needed for activation of ERK (AND gate) or either of them is sufficient for ERK activation (OR gate). In our graphical formalism, AND gates with multiple inputs are depicted as hyperedges (a hyperedge is a generalization of an edge that allows multiple inputs and outputs; in our case all hyperedges have only one output, see section Materials and methods). For example, a two-input AND gate for Raf + NFκB→ERK is depicted as a 'Y' shape upstream of ERK (Figure 1E). An OR gate with m inputs is depicted in the hypergraph by m edges or hyperedges entering a node (in this case Raf→ERK OR NFκB→ERK). Any possible Boolean function for ERK can be represented as a combination of some of these three edges/hyperedges, and identifying the correct combination is the main goal of our approach. In the remainder of this paper, we consider simple edges to be included within the general term 'hyperedge.'

Approach to model optimization

How should we maximize the match between model and data without overfitting (the introduction of excessive complexity)? Principles such as the Akaike Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (Schwarz, 1978), and Minimal Description Length (MDL; Barron *et al*, 1998) have been developed to formalize the concept of an optimal model. Boolean models with a fixed structure have no degrees of freedom and metrics such as AIC and BIC are

obviously not applicable. MDL provides an applicable theory for investigating the balance between fit and complexity (which scales with size and the number of degrees of freedom), but practical implementation of MDL requires making assumptions about how to encode complexity and typically involves the introduction of a tunable parameter that balances fit and complexity (Zhao *et al*, 2006). We therefore chose to use a bipartite objective function common in reverse engineering (Bonneau *et al*, 2006; Nelander *et al*, 2008) and LASSO regression (Tibshirani, 1994) that balances fit and size using a tunable parameter chosen to maximize the predictive power of the model. Other objective functions can be applied in the future, if theory or practice suggests that this will improve the outcome.

Given these assumptions, training a Boolean superstructure against experimental data is an optimization problem in a search space defined by the hypercube $\Sigma=\{0,1\}^r$ where



candidate solutions (models) are encoded in vectors $P \in \Sigma$ and r is the number of hyperedges in the superstructure model. Each hyperedge in the hypergraph is assigned an index i in vector P, $i=1,\ldots,r$, such that $P_i=1$ when the hyperedge is included in the model and 0 when it is not. The objective function for optimization is based on the mean squared error (MSE) deviation between the data and model (Θ_f), and a second term that penalized increasing model size (Θ_S). Thus, for a set of data containing n_E data points collected for m readouts (the four measures of protein activity in the current case) at n time points under s experimental conditions, (combinations of ligands and small-molecule drugs in our case) we minimize

$$\Theta(P) = \Theta_f(P) + \alpha \cdot \Theta_s(P) \qquad (1)$$

where $\Theta_f(P) = \frac{1}{n_E} \sum_{k=1}^{s} \sum_{l=1}^{m} \sum_{t=1}^{n} (B^M_{k,l,t}(P) - B^E_{k,l,t})^2$ and $\Theta_s(P) = \frac{1}{v_e^S} \sum_{e=1}^{r} v_e P_e$ such that $B^M_{k,l,t}(P) \in \{0,1\}$ is the value (0 or 1) as predicted by computation of the model's logical steady state (Klamt *et al*, 2006) and $B^E_{k,l,t} \in [0,1)$ is the data value for readout l at time t under the kth experimental condition. In this paper, we consider only one time point t after stimulation. To compute the size penalty, each hyperedge in a given solution P is weighted by the number of starting (tail) nodes v_e so that an AND gate representing Raf (*AND*) NFκB→ERK carries the same penalty as Raf→ERK (OR) NFκB→ERK and twice the penalty of a simple edge. By imposing a size penalty during optimization we ensure that unnecessary and redundant gates are not included in the final model. The size penalty is normalized to the size of the complete superstructure $v_e^S = \sum_{e=1}^{r} v_e$ and weighted with the tunable parameter α, which is chosen to maximize predictivity. The variable (P) is implicit when Θ, Θ_S, and Θ_f are mentioned below.

Equation (1) can be optimized by exhaustive evaluation of all possible solutions for the toy model, but the search space Σ increases exponentially with the number of hyperedges r. For large models we, therefore, implemented a genetic search algorithm (see section Materials and methods). We also compressed the search space Σ, based on the concept of Sperner systems (Bollobas, 1986). This obviates the need to search over redundant combinations of hyperedges. For example, nodes X and Y can be connected to downstream node A with three possible hyperedges: X→A, Y→A, and (X AND Y)→A. However, the logical combination of X OR (X AND Y)→A has the same truth table as X→A, but is larger and will not appear in optimized models. Thus, it is unnecessary to consider models containing the X OR (X AND Y)→A logic. We therefore replace the full set of Boolean functions (all possible combinations of hyperedges) in the model superstructure with a reduced set containing only the smallest non-redundant combinations of hyperedges, which corresponds to a Sperner hypergraph (see section Materials and methods).

As an illustration we trained the Boolean superstructure of toy models against synthetic data by optimizing equation (1) for several values of α. Since the training data was binary, Θ_f simply corresponded to the average number of wrong predictions. As the toy model was not necessarily identifiable, more than one model P (possibly many) can have the same value Θ. For example, with $\alpha=0$, four models with perfect fits to synthetic data ($\Theta_f=0$) were recovered. One of them corresponded to the reference model and the others differed in having alternate logic upstream of ERK (shown in Figure 1F, insets). When model size was penalized to a modest degree ($0<\alpha<0.25$), the smallest of the four $\alpha=0$ models was recovered and it corresponded to the reference model (Figure 1F). When model size was further penalized ($0.25<\alpha<0.86$) a single model was recovered in which NFκB was no longer linked to upstream and downstream nodes, giving rise to one mismatch between simulation and data (Figure 1G). With $0.75<\alpha<1.54$ a yet smaller model having six mismatches was obtained and, finally, with $\alpha>1.54$ the size penalty overweighed fit and calibration returned an empty model with no hyperedges and all nodes in their default state (0 for all nodes but IκB and GSK3, which were set to 1). These results correspond to a Pareto frontier with a trade-off between model size and goodness of fit (Figure 1H and I). Overall, this exercise illustrates the ability of a CNO-based workflow to regenerate a reference Boolean model using synthetic data and a signed directed graph (Figure 2). It also illustrates the importance of having a rich data set. For example, omitting the combined treatment of TGFα and TNFα from the synthetic data prevents recovery of the AND gate that lies upstream of NFκB in the reference model.

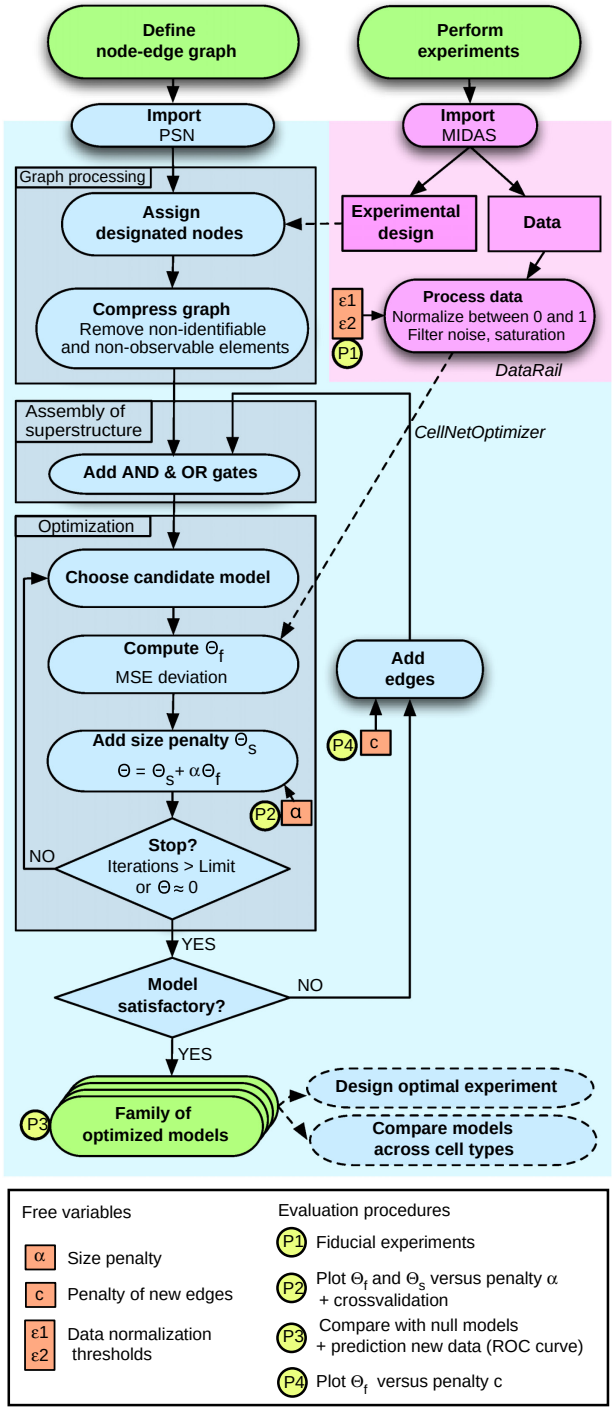
Applying CNO to growth and inflammatory signalling in human cancer cells

Having tested CNO on a toy network, we turned to the analysis of real data collected from human liver cancer cells. Hepatocytes, which constitute the majority of cells in the liver, are both targets for and sources of multiple chemokines and cytokines that activate overlapping intracellular signalling pathways. To begin to understand the cooperative and antagonistic interactions among ligands, we used a cue–signal–response (CSR) data set from HepG2 hepatocellular carcinoma cells exposed to one of seven cytokines in the

Figure 1 Assembly, calibration, and analysis of a toy signalling model. (**A**) Signed directed graph representing a simple pathway as visualized using Cytoscape (Shannon *et al*, 2003). The topology of the reactions downstream of TGFα and TNFα receptors is imaginary, but it includes real molecules such as Shc, Ras, Raf, MEK, ERK, PI3K, AKT, GSK3, IκK, IκB, NFκB, TRADD, caspase 8 (denoted C8), and the Grb–Sos complex (denoted GrbSos). (**B**) The design of the synthetic experiments used to train the graph in panel A. Each column represents an experiment and each row a different designated species as follows: green denotes ligands, red denotes the protein targets of kinase inhibitors, and blue denotes the proteins whose states were assayed (readouts). The presence or absence of ligand or an inhibitor specific to a node is denoted with '+' and '−', respectively. The 0/1 value for the readouts corresponds to the result obtained from simulating the reference model under specific conditions of ligand and inhibitor exposure. (**C**) Rules applied to graphs to create compressed representations. (**D**) The experimental design (B) determines which nodes in the graph are designated and which are undesignated. This information, in combination with the rules in panel C was used to create a compressed graph, with nodes eliminated by compression indicated by dashed lines. (**E**) Superstructure of all models compatible with the graph in panel A. (**F**) Optimal models for size penalties of $0 \leqslant \alpha \leqslant 0.23$. The highlighted panels to the right (boxed with dashed lines) show three different logical structures recovered during model calibration with $\alpha=0$. The fit to data was perfect for all models ($\Theta_f=0$). (**G**) Optimal model for $0.23 \leqslant \alpha \leqslant 0.75$. The matrix below shows the single mismatch (in red) between model-based simulations and the training data. (**H**, **I**) Balance between the fit of the data Θ_f (the MSE deviation from data; see text for details) and size Θ_S for models recovered using different values of the size penalty, α.

presence or absence of seven small-molecule kinase inhibitors, and then measured the states of 16 intracellular signalling proteins before and 30 min after ligand addition (Figure 4A). The extracellular ligand 'cues' included two mediators of the acute-phase response, TNF-α and interleukin-6 (IL-6); the TLR4 agonist lipopolysaccharide (LPS); two general inflammatory factors active in liver, IL-1α and interferon-γ (IFN-γ); and two mitogenic factors, insulin-like growth factor-1 (IGF-1) and the EGFR ligand, transforming growth factor-α (TGFα; Supplementary Table 1). Cells were exposed to one of seven small-molecule kinase inhibitors at concentrations sufficient to achieve ~90% target inhibition (as assaying in HepG2 cells, Alexopoulos *et al*, in preparation). After1 h, cells were treated with ligands and samples were then collected at 0 and 30 min, and the phosphorylation of 16 intracellular signals was measured in whole-cell lysates using bead-based sandwich ELISA methods (multi-analyte profiling xMAP technology; Luminex, Austin, TX, USA). Further rationale for the experimental design can be found in reference Alexopoulos *et al* (in preparation), but with respect to our goal of training a Boolean model, the ~1000 biochemical measurements in the data set represent a relatively rich set collected from cells under different conditions (combinations of ligands and inhibitors).

We constructed a signed directed graph of intracellular signalling covering the ligands and immediate-early kinase pathways in our data set using the software from Ingenuity Systems (http://www.ingenuity.com/). The graph was supplemented with literature data on IRS-1, whose representation in the Ingenuity database seemed particularly poor (see section Materials and methods). The resulting literature-derived protein signalling network (LD-PSN) contained 82 nodes and 116 edges comprising 26 designated and 56 undesignated nodes. Compression with CNO simplified the graph to 31 nodes and 53 edges. Creating all logical combinations consistent with this compressed graph yielded a superstructure with 131 hyperedges. In the absence of compression the superstructure would have contained 197 hyperedges.



Normalizing experimental data

Variables in Boolean models are necessarily binary (0 or 1), but the biochemical measurements in our CSR data set are continuous. The simplest way to compare the experimental data to model output is to discretize the data to a value of 0 or 1 based on a set of thresholds. However, discretization reduces

Figure 2 Workflow for creating and calibrating Boolean models using CNO software. Signed directed graphs are imported into CNO (a set of software routines implemented in MATLAB) and data are imported into DataRail (Saez-Rodriguez *et al*, 2008). The experimental design defines which nodes in the graph are designated and which are undesignated. The graph is then compressed based on three procedures operating on undesignated nodes (see Figure 1C). The compressed graph is transformed into a superstructure that represents a superposition of all Boolean models compatible with the graph. An optimization algorithm then searches the superstructure for those models that minimize the value of the objective function Θ for a specific value of the size penalty, α; typically this calibration procedure is repeated for multiple values of α (see text for details). Optimization is terminated when a predetermined criterion is fulfilled; typically the number of times optimization is performed or when a threshold value for Θ is reached. Optimization can then be terminated or new routines initiated to add new edges to the optimized model, followed by another round of calibration aimed at decreasing the value of the objective function. During edge addition, a higher size penalty (c_i) is assigned to edges absent from the initial graph to reflect the fact they are not supported by prior knowledge. Once a model has been found, different types of analyses can be performed, such as designing new experiments based on model predictions or comparing models between different cell types. Moreover, a series of evaluation procedures should be performed that include cross-validation, ROC curve, and comparison to null models (indicated in yellow; see text for details).

the information content of the data by implying the existence of unrealistic on-off signalling states and is not necessary: the MSE deviation can be computed by comparing binary model outputs to normalized continuous data (i.e., values between 0 and 1). Normalization commonly involves dividing each measurement in a series by the highest value, but this over-emphasizes outliers while underweighting small but highly reproducible differences. We therefore developed a multi-step scheme for non-linear data normalization (see Figure 3). For each readout l, we specified a lower limit based on experimental noise ($S_{l,N}$) and an upper limit corresponding to saturation of the assay ($S_{l,\mathrm{SAT}}$). In the current work, analysis of serially diluted samples showed $S_{l,N}\sim 500$ and $S_{l,\mathrm{SAT}}\sim 18\,000$. Next, using routines newly added to the DataRail software (Saez-Rodriguez *et al*, 2008), we computed the ratio $S^{\mathrm{R}}_{k,l,t}$ (where k is the index for the experimental condition, l for the measurement—the specific xMAP sandwich immunoassay in this case—and t is the time point) between each measurement, $S_{k,l,t}$, and $S_{k,l,0}$, the value at the start of the experiment, as follows: $S^{\mathrm{R}}_{k,l,t}=|(S_{k,l,t}-S_{k,l,0})|/S_{k,l,0}$ (Figure 3 illustrates how this methods applies for signals that rise or fall after $t=0$). In our data, $S^{\mathrm{R}}_{k,l,t}$ varied between 0 and 80, and was mapped to a value $0<V^{\mathrm{R}}_{k,l,t}<1$ using a sigmoidal normalization function that maximized sensitivity in the intermediate range. The parameter ε_1 defines the midpoint of the normalization function (i.e., when $V^{\mathrm{R}}_{k,l,t}=0.5$) and was constant across the entire data set. A value for ε_1 was chosen heuristically based on the subset of CSR data for which we had strong prior expectations; we refer to these as the 'fiducial data'. For example, we know that treatment of cells with TGFα (an EGFR ligand) triggers a dramatic increase in ERK phosphorylation except when the MEK kinase inhibitor PD325901 is present (Figure 4A). Similarly, TGFα triggers AKT phosphorylation except when PI3K is inhibited by the small-molecule inhibitor ZSTK474. We therefore chose a value for ε_1 such that $V^{\mathrm{R}}_{\mathrm{ERK},l,t}>0.5$ in cells treated with TGFα and $V^{\mathrm{R}}_{\mathrm{ERK},l,t}<0.5$ in cells treated with TGFα plus PD325901, while simultaneously yielding $V^{\mathrm{R}}_{\mathrm{AKT},l,t}>0.5$ in cells treated with TGFα and $V^{\mathrm{R}}_{\mathrm{AKT},l,t}<0.5$ in cells treated with TGFα and ZSTK474. Overall, such fiducial experiments comprised 5% of the total data set. In the future, we anticipate implementing a scheme for ε_1 optimization based on multiple user-defined fiducial data points. However, data normalization in the current work was not very sensitive to the precise value of ε_1 (see below) and a value $\varepsilon_1=0.5$ proved effective.

A subtlety in data processing is that it is necessary to remove from the data phospho-protein measurements for nodes whose activities are blocked with a drug. For example, removal of phospho-MEK data collected from cells treated with the MEK inhibitor PD325901, because phosphorylation levels are not a reliable measure of the activity of PD325901-bound kinase. Moreover, to deemphasize data in which measurements were close to background for all time points under condition k, we computed the ratio of each data point $S_{k,l,t}$ to the maximal value obtained for the same readout l across all conditions and time points in the full data set ($S_{l,\mathrm{MAX}}$), and then transformed this into a value $0<V^{\mathrm{M}}_{k,l,t}<1$ using a saturation curve (Langmuir function, which has the same form as the Michaelis–Menten function) with half-maximal value at $\varepsilon_2\sim 0.05$. The parameter $V^{\mathrm{M}}_{k,l,t}$ was then used to penalize very weak signals by computing the product $B^{\mathrm{E}}_{k,l,t}=V^{\mathrm{M}}_{k,l,t}\cdot V^{\mathrm{R}}_{k,l,t}$. Finally, the MSE deviation Θ_f between the experimental value $B^{\mathrm{E}}_{k,l,t}$ and its corresponding simulated value $B^{\mathrm{M}}_{k,l,t}$ was computed as a measure of the fit of model to data. In cases in which phosphorylation corresponds to repression of a node (in our case this is true for GSK3 and IκB) the appropriate simulated value for computing MSE deviation is $1-B^{\mathrm{E}}_{k,l,t}$.

As data are continuous and Boolean models are binary, a residual 'discretization' error remains even in the case

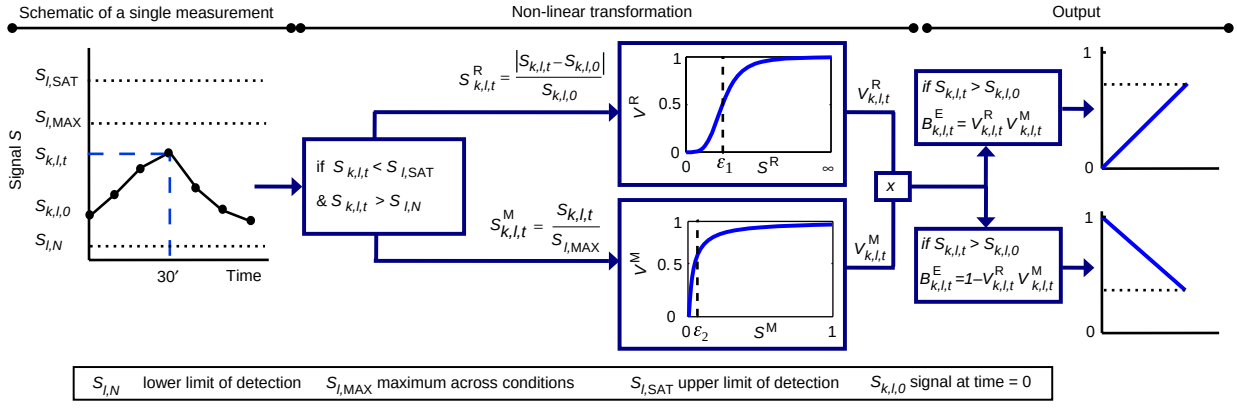
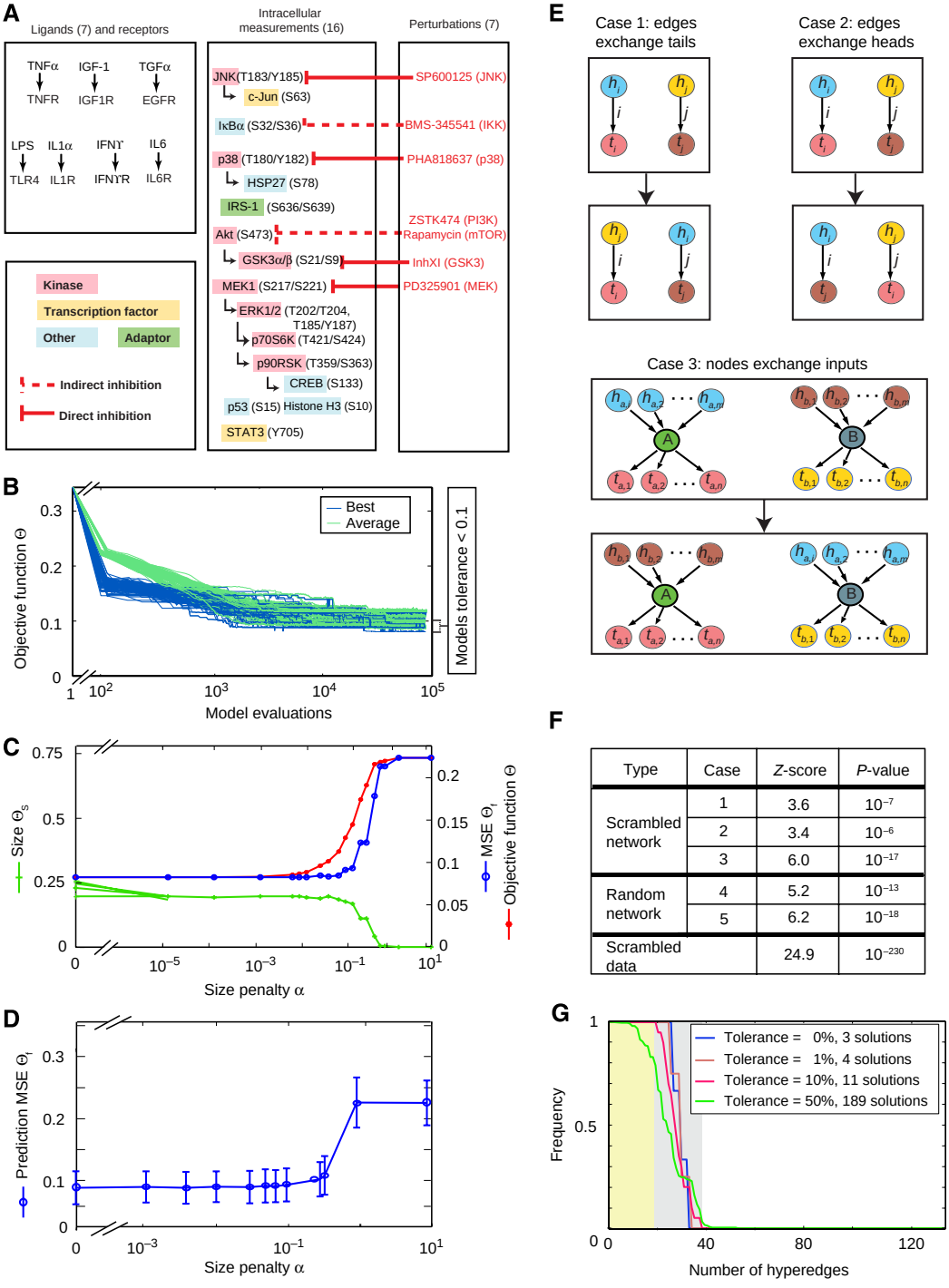


Figure 3 Procedure for data normalization. If the measured signal $S_{k,l,t}$ for readout l at time t under the kth experimental condition is either above the saturation limit ($S_{l,\mathrm{SAT}}$) or below the limit of detection ($S_{l,N}$) of the lth measurement method, the value is not reliable and is therefore ignored; values for $S_{l,\mathrm{SAT}}$ and $S_{l,N}$ are obtained from serial dilution experiments. Otherwise, the scaled measurements are computed relative to the value of the measurement at the start of the experiment $S^{\mathrm{R}}_{k,l,t}=|(S_{k,l,t}-S_{k,l,0})|/S_{k,l,0}$ and transformed using a non-linear normalization function (Hill function; upper part of the schematic) into a value $0<V^{\mathrm{R}}_{k,l,t}<1$. To impose a penalty on measured values that are very low relative to other time points and experimental conditions, the value is scaled relative to the maximum ($S^{\mathrm{M}}_{k,l,t}=S_{k,l,t}/S_{l,\mathrm{MAX}}$) and transformed $0<V^{\mathrm{M}}_{k,l,t}<1$ using a saturation curve (e.g., Langmuir function; lower part of the schematic). Values for adjustable parameters ε_1 and ε_2 specifying midpoints of the data normalization functions are determined from a 'fiducial' subset of data as described in the text. The two-scaled and normalized values for each data point are then multiplied, $B^{\mathrm{E}}_{k,l,t}=V^{\mathrm{M}}_{k,l,t}\cdot V^{\mathrm{R}}_{k,l,t}$, to yield the value used for model calibration. Calibration involves minimizing the MSE deviation between all experimental measurements $B^{\mathrm{E}}_{k,l,t}$ and model outputs $B^{\mathrm{M}}_{k,l,t}$. The data normalization procedure is embedded in DataRail and is a generalization of the discretization algorithm described by Saez-Rodriguez *et al* (2008).

of the best fits. This residual error Θ_f^{D} corresponds to the difference between the discrete predictions of the best possible Boolean model and the continuous data: $\Theta_f^{\mathrm{D}} = \frac{1}{n_{\mathrm{E}}} \sum_{l=1}^{s} \sum_{k=1}^{m} \sum_{t=1}^{n} (B_{k,l,t}^{\mathrm{D}} - B_{k,l,t}^{\mathrm{E}})^2$, where $B_{k,l,t}^{\mathrm{D}}$ is the binary value arising from rounding $B_{k,l,t}^{\mathrm{E}}$. With the optimized model and CSR data set in this paper, Θ_f^{D}=0.024. Further investigation of data normalization procedures is no doubt warranted, ideally based on maximizing the predictivity of trained models. However, we observed that varying ε_1 from 0.3 and 0.7 did not change the set of optimal models, although it did alter identifiability (see below). We therefore concluded that our approach is not unduly sensitive to changes in the values of adjustable parameters used for data normalization.



Type	Case	*Z*-score	*P*-value
Scrambled network	1	3.6	10^{-7}
	2	3.4	10^{-6}
	3	6.0	10^{-17}
Random network	4	5.2	10^{-13}
	5	6.2	10^{-18}
Scrambled data		24.9	10^{-230}

Training a multi-receptor model against experimental data

The Boolean superstructure for the seven-receptor network (containing 131 hyperedges) was calibrated against normalized biochemical data (comprising 809 data points) by running a genetic algorithm multiple times and monitoring the objective function to ensure stability of the solution at the end of each run (Figure 4B). Even with α=0 we found that optimized models were about one-third the size of the initial superstructure (which contained all possible logical models), but exhibited fourfold improved goodness of fit to data. Further analysis revealed that the superstructure predicted many responses that were absent from the data because the LD-PSN contained too many interactions. As a consequence, an empty model—one containing 31 nodes but no hyperedges–actually had a lower value of Θ_f than the superstructure. To explore the relationship between Θ, Θ_f, and Θ_S we performed 20 rounds of optimization at each of 19 values of α between 0 and 50. Over this range, the size of optimized models was nearly constant at $\sim$0.19, as was the goodness of fit, until $\alpha > 0.1$ at which point the model collapsed and the fit approached that of the empty model (Figure 4C). This suggests a penalty of $0 < \alpha < 0.1$. To confirm this range, we performed 10-fold cross-validation by constructing models from 90% of the data and then attempted to predict the remaining 10%. The process was then iterated $\sim$150 times for different values of α. Trained models predicted missing data most accurately with $\alpha < 0.1$ (Figure 4D). We therefore concluded that over the range $0 < \alpha < 0.1$ calibrated models have a complexity close to the minimum value necessary for a good fit, and we chose a value α=0.0001 for the remaining of the analysis.

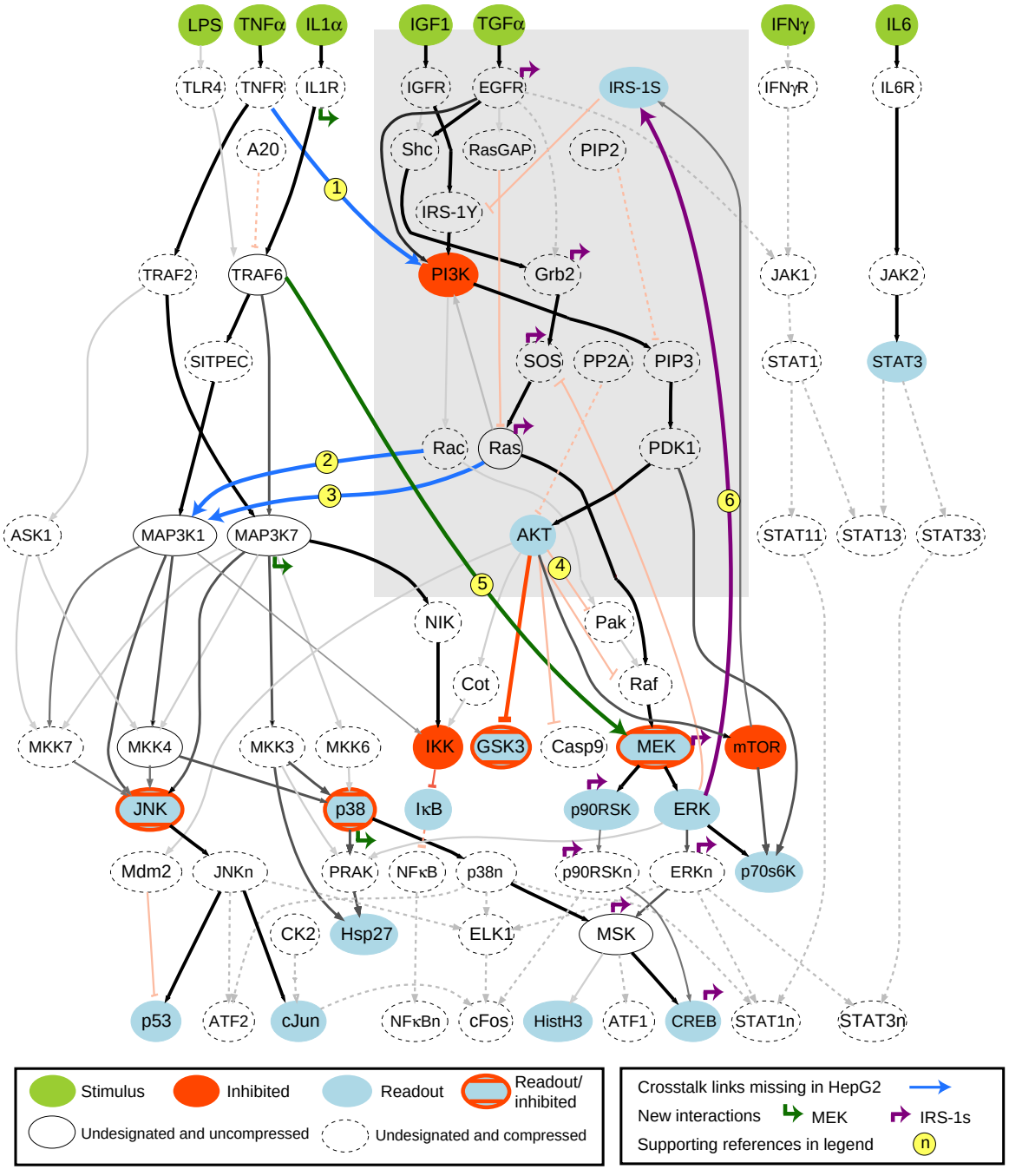
Statistical significance of trained models

Are model topologies recovered by calibration statistically significant given the training data and the prior knowledge in the LD-PSN, or might they arise by chance? To address this question we generated scrambled versions of the data and both scrambled and random model superstructures. First, 500 sets of scrambled data were generated by pairwise exchange of data points. This was accomplished by randomly dividing the original data set in two and swapping all data points, thereby scrambling relationships between signals and experimental conditions. For each of the 500 scrambled data sets, we repeated calibration and observed that the fit of optimized models to data was significantly worse than the fit of calibrated models to unscrambled data ($P < 10^{-200}$). Next, 500 scrambled networks (null networks) were created from the LD-PSN superstructure by random pairwise exchange of elements. We performed three types of exchanges (see Figure 4E): in type-1 the tails for two hyperedges chosen at random were exchanged and the process was iterated across all hyperedges (keeping constant the out-degrees of all nodes); in type-2 this procedure was followed for the heads of pairs of hyperedges (keeping constant the in-degrees of all nodes); and in type-3 all of the edges ending at each of two randomly chosen nodes were exchanged and the process was iterated across all nodes (keeping constant the distribution of in-degrees of the nodes in the graph). Alternatively we created a set of completely random networks having (i) the same number of nodes and edges as the LD-PSN, (ii) at least one edge per node, and (iii) network inputs (corresponding to cytokine stimuli in the LD-PSN) with no incoming edge but at least one outgoing edge. Two types of random networks having these characteristics were generated: type-4 in which all measured nodes had at least one input, as in the LD-PSN and type-5 in which this restriction was removed. For 500 runs over superstructures derived from each of the five types of null network, the LD-PSN superstructure yielded a significantly better fit to data: for null model type-1 and 2, the P-values were $\sim 2 \times 10^{-6}$ to 3×10^{-7}, whereas random networks type-4 and 5 yielded P-value $\sim 10^{-13}$ and 10^{-18} (Figure 4F). The discrimination between real and null networks of type-1 and 2 was less than that for random networks because the former retain information on hub nodes (e.g., Ras). In type-3, where inputs but not outputs were scrambled, the null networks were no better than random networks (P-value $\sim 10^{-17}$). From these data we conclude that the LD-PSN contains prior knowledge, with a significantly greater match to experimental data than would occur by chance.

Figure 4 The selection of models of HepG2 hepatocellular carcinoma cells. (**A**) Design of experiments in the training data set depicting the use of seven ligands and seven drugs. The 16 proteins that were measured are depicted by coloured boxes with specific states of modification (shown in parentheses) that were assayed by xMAP technology. The red inhibitory arrows depict the seven small-molecule drugs that were used to block protein kinases; drugs were added to HepG2 cells at concentrations two- to four-fold above their measured IC50s 60 min prior to ligand addition (see Alexopoulos *et al*, in preparation, for details). (**B**) Evolution of model calibration with a genetic algorithm run 100 times. For each run, a set of 100 models (chosen at random from all possible models compatible with the superstructure) was analysed. At each generation in the genetic algorithm, the average (green) and best (blue) value of Θ across the set of 100 models was evaluated. Sufficient numbers of evaluations were performed ($\sim 10^5$) to obtain stable solutions. (**C**) Trade-off between fit of data and size of model. The total objective function Θ (red line), MSE deviation of model from data Θ_f (blue line), and model size Θ_S (green line) are shown for the best model recovered at 19 different values of the size penalty, α. With α=0, multiple solutions were recovered having an equal value of Θ_f but different values of Θ_S; this is depicted in the figure by multiple converging green lines. (**D**) Predictive power as estimated by 10-fold cross-validation. For each value of α, $\sim$150 trainings were performed, leaving out one-tenth of the data randomly chosen. The plot shows the mean and standard deviation of the prediction of the left-out data. Predictive power is best for $\alpha < 0.1$. (**E**) Different approaches used to scramble prior knowledge encoded in the LD-PSN. In scrambled networks of type-1, edges are divided into two random and equally sized groups; edges in the first group exchange their head (input node) with edges in the second group; the process is iterated over all nodes. Type-2 of scrambling is equivalent, but edges exchange tails (output nodes). In type-3, nodes are divided in two random groups and exchange all incoming edges as a group. (**F**) Statistical evaluation of the training of the randomized and scrambled models to the real data, and of the Ingenuity network to the scrambled data. (**G**) Distribution of hyperedges across families of calibrated models as a measure of model identifiability. Each curve represents a sorted histogram depicting the frequency with which a hyperedge was recovered based on the allowable tolerance between models under consideration and the best model where tolerance is defined as the increase in Θ_f relative to the lowest value achieved (Θ_f=0.081) as follows: blue line, 0% tolerance—3 models; brown line, 1% tolerance ($0.081 \leqslant \Theta_f < 0.083$)—4 models; red line, 10% tolerance ($0.081 \leqslant \Theta_f < 0.089$)—11 models; and green line, 50% tolerance ($0.081 \leqslant \Theta_f < 0.122$)—189 models. For the 11 best models (10% tolerance), a yellow band denotes hyperedges present in all models, the grey band hyperedges present in some models, and the white band hyperedges absent from all models.

Features of trained models

With α=0.0001 and ~300 rounds of optimization, we obtained three best-fit models that had Θ_f=0.081, a value that was roughly three time the residual error of Θ_f^D=0.024, showing that even the best model did not fit data perfectly (Θ_f could be reduced further by adding new interactions; see below). The recovery of multiple solutions with the same value for Θ shows that the model is not completely identifiable. We therefore divided the number of occurrences of a particular hyperedge by the number of calibration runs; this value is an estimator of the probability, p_i, that the ith hyperedge is actually present in the true model (i.e., $p(\text{hyperedge}_i|\text{data})$). In the case of complete identifiability, p_i is either 1 or 0 for all hyperedges, and standard deviation $\sigma_i=\sqrt{(p_i\cdot(1-p_i))}=0$. Complete non-identifiability corresponds to p_i=0.5 and σ_i=0.5 for all hyperedges. As expected, identifiability varied with the allowable deviation from the lowest value of Θ_f achieved. In principle this deviation should be similar to the propagated error in the training data. A 10% error in phospho-protein measurements (close to the error we obtain upon repeatedly assaying the same biological samples; data not shown) when propagated through our normalization



procedures yields ~4% variance in $B^{E}_{k,l,t}$ and thus a 2% tolerance in MSE. However, if we allow a more conservative 10% tolerance in goodness of fit to allow for biological variability in the experiment so that $\Theta_f < 0.089$, 11 models were recovered and $<\sigma \geqslant 0.05$. With a tolerance of 50%, 189 models were included with $\Theta_f < 0.122$ and $<\sigma \geqslant 0.09$.

The set of 11 calibrated models had 26–28 hyperedges of which 19 were present in all models (Figure 4G, yellow band). Seventeen (grey band) of the 131 hyperedges in the original superstructure (white band) were present in some but not all models, whereas 95 were absent from all models. In Figure 5 estimates of p_i for 11 models and $<\sigma \geqslant 0.05$ are shown by line weights: thick black lines denote hyperedges that were preset in all models, grey lines denote hyperedges that were present in some models, and light grey lines denote hyperedges in the superstructure that were absent from all calibrated models. From this representation we can see that, as expected, non-identifiable hyperedges involved proteins that were neither perturbed nor assayed; for example, the multiplicity of MAP kinase-kinases that regulate p38 and JNK are largely indistinguishable. In future experiments it should be possible to increase model identifiability by adding additional data on the phospho-states of individual MAPKKs.

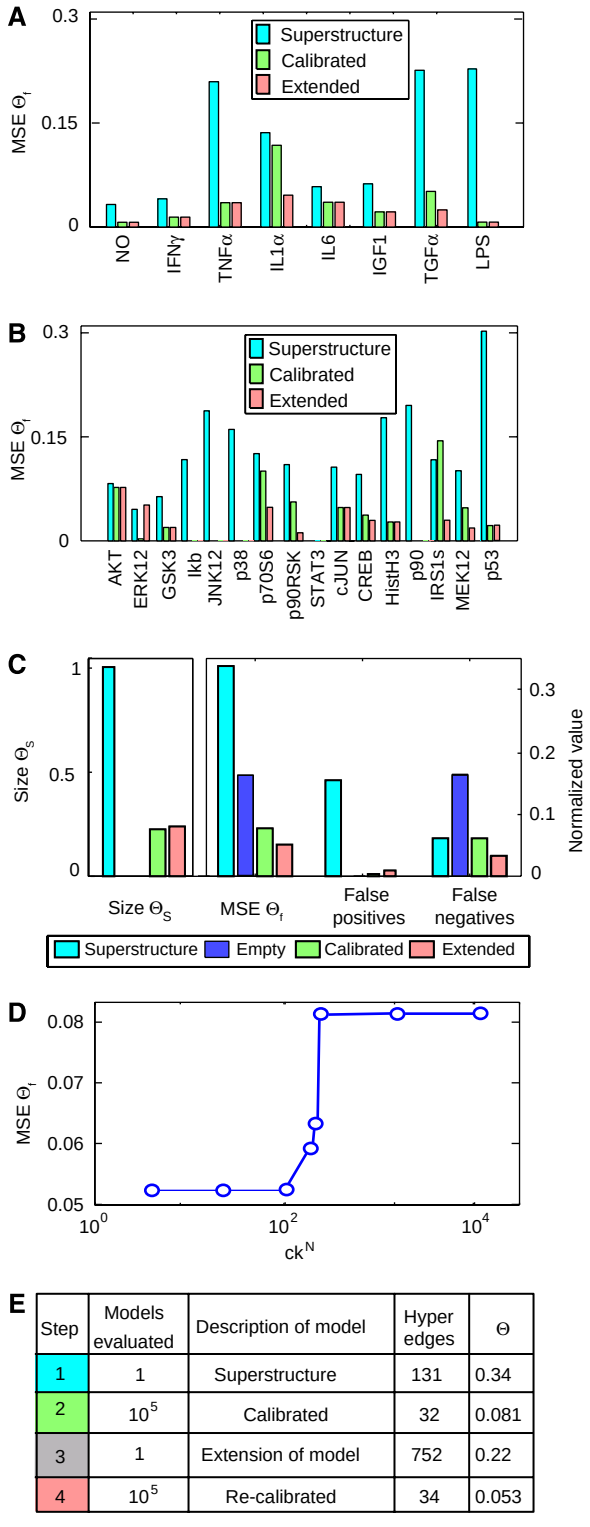
In optimized models, all receptors are linked to downstream signalling molecules, with the exception of TLR4. This is not a spurious result since models calibrated to data from other cell types contain links between TLR4 and downstream signalling proteins (data not shown). Instead, TLR4 receptors in HepG2 cells do not appear to be active (Alexopoulos *et al*, in preparation). A surprising feature of optimized HepG2 models is the relative paucity of links between intracellular molecules activated by inflammatory factors and those activated by growth factors. Specifically, the link TNFR→PI3K (labelled 1; Figure 5) proposed by Marchetti *et al* (2004) was missing as was Rac→MAP3K1 (labelled 2; Fanger *et al*, 1997) and Ras→MAP3K1 (labelled 3; Russell *et al*, 1995). Crosstalk between AKT and the Raf/MEK/ERK cascade (labelled 4; Guan *et al*, 2000) was also missing. The absence of these interactions from the LD-PSN calibrated to data from HepG2 cells does not appear to be an artefact of our approach because links 1 and 3 were present in a preliminary model assembled using data from Huh7 cells, another hepatocellular carcinoma cell line (data not shown). Thus, we propose that the exclusion of documented protein–protein interactions from calibrated models reflects their irrelevance in HepG2 cells within the first 30 min after ligand addition.

Identifying new interactions that improve fit to data

Although the LD-PSN-derived superstructure contained considerably more hyperedges than were present in the calibrated models, it nonetheless seemed likely that some interactions might be missing, either because the literature survey was imperfect or because our understanding of the relevant biology is incomplete. We therefore asked whether addition of a limited number of hyperedges would improve fit to data. The number of possible edges in a graph of n nodes is $2(n^2-n)$ (because each edge can point either direction and be either positive or negative), and the number of hyperedges increases as $n(3^{n-1}-1)$. Thus, a LD-PSN with 82 nodes has 13 284 possible edges and the associated superstructure has $\sim 10^{40}$ possible hyperedges. Even the compressed superstructure of 31 nodes has 1860 potential edges and $\sim 10^{15}$ hyperedges. Thus, the LD-PSN superstructure contained ~1% of all possible edges for a graph of 82 nodes. The search space for new edges scales as 2^y where y is the number of hyperedges, making it impossible to perform an exhaustive search. We therefore focused our search on areas of the model in which the fit to data was poor (CNO ranks stimuli, perturbations, and readouts according to Θ_f). In our data, the greatest deviation between the nine best-fit models and data involved IL1α and TGFα stimulation (Figure 6A), and assays for IRS-1S and p70S6K phosphorylation (Figure 6B). Accordingly, 630 OR-gated hyperedges were added to the LD-PSN-derived superstructure to connect nodes downstream of IL1aα and TGFα to IRS-1S and p70S6K, or to nodes upstream of IRS-1S and p70S6K. Only positive (activating) hyperedges were evaluated, since errors involved mostly false negatives (Figure 6C). The search for new hyperedges involving these nodes was accomplished using a modified size penalty $\Theta^*_s = \sum_{k=1}^{n} c_k v_k P_k$

Figure 5 Family of calibrated models recovered for immediate-early signalling downstream of seven transmembrane receptors in HepG2 cells. The graph represents the nodes and edges present in a set of calibrated Boolean models considered to be indistinguishable based on the data (see text for details). The graph was created using a routine in CNO based on a GraphViz visualization engine (www.graphviz.org), followed by manual annotation using Adobe Illustrator. Green ellipses denote stimuli, red ellipses species blocked by kinase inhibitors, and blue ellipses denote readouts. Ellipses with red borders and blue filling were both measured and subjected to inhibition using small-molecule drugs. Ellipses with dashed borders were compressed during graph processing, and empty ellipses were not designated but were not compressed since they did not fulfill the three rules for compression (Figure 1C). Positive interactions are denoted in grey and black and inhibitory interactions in red. We did not recover any AND gate; this is, however, not an artefact of the model, but rather a feature extracted from the data (in other data sets using data from other cell types, calibrated models contain AND gates). Colour and line thickness denote the frequency with which each hyperedge was present in the models; hyperedges represented by solid black lines were present in all models, grey hyperedges were present in some models, and dashed grey (activating) hyperedges or dashed red (inhibitory) hyperedges were absent from all the models. CNO automatically determined that none of the proteins downstream of IFN-γ were assayed or inhibited and thus, this input remains isolated; model decompression introduces the possible connections making it possible to visualize the calibrated model in the context of prior knowledge. Blue arrows highlight key interactions present in the starting, literature-derived PSN, but excluded from calibrated models, and connecting growth factor signalling pathways downstream of IGF-1 and TGFα (shadowed in grey) and inflammatory pathways. The existence of these interactions has been documented in the literature (with numbering indicated by the yellow circles) (1) by Marchetti *et al* (2004); (2) by Fanger *et al* (1997), and (3) by Russell *et al* (1995). Moreover, in a preliminary model using data from Huh7, another liver cancer cell line, interactions (1) and (3) were present. Crosstalk between AKT and ERK (4), described by Guan *et al* (2000) was not observed in models of either HepG2 or Huh7 cells. Green and purple arrows denote hyperedges that impinge on either MEK or phospho IRS-1(S636/S639) that were absent from the LD-PSN but were identified during model extension as improving fit to data. The short arrowheads depict alternative origins for links that are indistinguishable because of model non-identifiability: each model would contain only one green and one purple link. TRAF6 has been reported to be an upstream regulator of MEK (interaction 5) by Hers and Tavaré (2005), and support for the role of ERK in the phosphorylation of some serine residues in IRS-1 (interaction 6) can be found in reference Rhee *et al* (2004).

that included a weighting factor c_k that varied from hyperedge to hyperedge. This made it possible to penalize newly added hyperedges more heavily than hyperedges derived from prior knowledge. We assigned a value c_k=1 for existing hyperedges and investigated the effect of assigning different values of c_k to the new hyperedges (c_k^N). The optimal penalty for new hyperedges is one that results in new links only if they provide a significance improvement in the fit to data. Model calibration was performed using the extended superstructure and different values of c_k. With $c_k^N \geqslant 500$ no new hyperedges were added, since the increase in size Θ_s* outweighed the decrease in MSE error Θ_f (Figure 6D), but, for $c_k^N \leqslant 100$, lower values of Θ_f. were obtained. With $c_k^N \ll 100$ Θ_f did not improve further, but there was the risk that links from the LD-PSN were replaced with alternatives for which there was no prior knowledge. Thus, c_k^N=100 appeared to be a near-optimal value for penalizing new edges.



Step	Models evaluated	Description of model	Hyper edges	Θ
1	1	Superstructure	131	0.34
2	10^5	Calibrated	32	0.081
3	1	Extension of model	752	0.22
4	10^5	Re-calibrated	34	0.053

The search for new interactions recovered a family of models in which Θ_f=0.053, a lower value than the previous best fit of Θ_f=0.081 (Figure 6E). Since Θ_f^D=0.024, this represents an improvement in fit of ~50% and was associated with an increase in the true positive rate (see below). The new models contained two new edges each, but the starting and ending points of the edges varied with the solution (due to non-identifiability). One set of edges linked linking nodes between IL1R and p38 to MEK (green arrowheads in Figure 5) and another linked nodes between EGFR and ERK to phospho-serine IRS-1 (IRS-1S; purple arrowheads). Although absent from LD-PSN, we nonetheless found literature support for a connection between TRAF6 and MEK in non-transformed human colonic epithelial cells (corresponding to the green line in Figure 5, labelled 5; Rhee *et al*, 2004), and for a connection between ERK and IRS-1 phospho-serine in primary adipocytes (Figure 5, magenta line, labelled 6; Hers and Tavaré, 2005). Hers and Tavaré (2005) also report the absence of a link between mTOR and IRS-1 phospho-serine, in agreement with our models, but different from what has been observed for other cell lines (Ozes *et al*, 2001; Hers and Tavaré, 2005).

Model validation

To evaluate the performance of optimized Boolean models we asked how well they could predict a 'validation data set' that was distinct from the training set. New experiments were performed in which HepG2 cells were treated with combinations of two ligands (IL6 + IL1α or IGF-1 + TGFα) in the presence and absence of small-molecule inhibitors of p38, MEK, PI3K, and EGFR (gefitinib; Herbst, 2002) protein kinases

Figure 6 Summary statistics for Boolean modelling of HepG2 signalling. (**A**) Average deviation from data for the untrained model superstructure (blue bars), best-fit calibrated model (green bars), and extended model having two added hyperedges (pink bars) sorted by ligand. (**B**) Average deviation as in panel A but sorted by intracellular signalling protein. (**C**) Size and fit to data during model assembly and optimization starting with full superstructure (blue), empty model (dark blue), calibrated model (green), and extended model containing two new hyperedges (pink). For simplicity, only one model of the family of solutions is represented. The left panels depict model size (Θ_S; left vertical axis) the and right panel shows the normalized number of false positives, false negatives, and the MSE deviation from data (right vertical axis). False-positive values arise when the model incorrectly predicts induction of signal and false-negative values when the model does not predict induction of signal that is found in the data. The empty model has no hyperedges and thus all states but IκB are zero. (**D**) MSE error Θ_f recovered upon training of the extended network with different values of the weight for the new hyperedges (c_k^N). (**E**) Computational cost of successive steps in model assembly and calibration, and the average value of the objective function achieved.

(Figure 7A). Phospho-protein measurements were made prior to and 30 min after addition of exogenous ligand, as in the training data. Sixty of the 88 conditions (77%) in the new validation data set were different from those in the training data set; overlap of the remaining 23% of the data allowed us to control for experimental reproducibility. For simplicity, the validation data was compared to simulations based on a single calibrated model of the family of solutions. The deviation between this model and the validation data was Θ_f=0.096, with a residual error Θ_f^D=0.03. This fit was nearly as good as the fit of the calibrated model to the training data (Θ_f=0.081 and Θ_f^D=0.024) and much better than the fit of the LD-PSN superstructure (Θ_f=0.28). Moreover, when the two data sets were combined into a single training set, the structure of the re-optimized models did not change, although identifiability improved slightly ($<\sigma>$ decreased from 0.14 to 0.13).

To measure the predictive power of best-fit models, we determined the rate of false-positive and false-negative predictions. False positives corresponded to cases in which the model predicted induction of signal (state=1), but the normalized experimental value was below 0.5. Analogously, false negatives corresponded to the cases in which the model did not predict induction of signal (state=0), but the normalized data was above 0.5. Both the binary rate (based on a simple count of errors) and a weighted rate, in which each error was multiplied by $(B^E_{k,l,t}-B^M_{k,l,t})^2$, were computed. The weighted rate is less familiar but probably better since errors are scaled by their magnitude (Figure 7D). Receiver operating characteristic (ROC) curves using either error metric show that the solution with α=0.0001 (the value we chose based on the cross-validation studies; Figure 4) exhibits the best ratio of false negatives (0.33 and 0.28 for the binary and weighted rates, respectively) to false positives (0.037 for the binary and 0.024 for the weighted rates). Values of α higher than 0.1 marginally decreased the ratio of false positives, but at the price of a significant increase in false negatives (1−true positives). Furthermore, inclusion of two new interactions (red data points in Figure 7D) improved the false-negative ratio (from 0.33 to 0.21, and 0.28 to 0.15 for the binary and weighted rates, respectively), with a modest increase in false positives (0.031 and 0.049). From these data we conclude that our optimized model has good predictive power.

Discussion

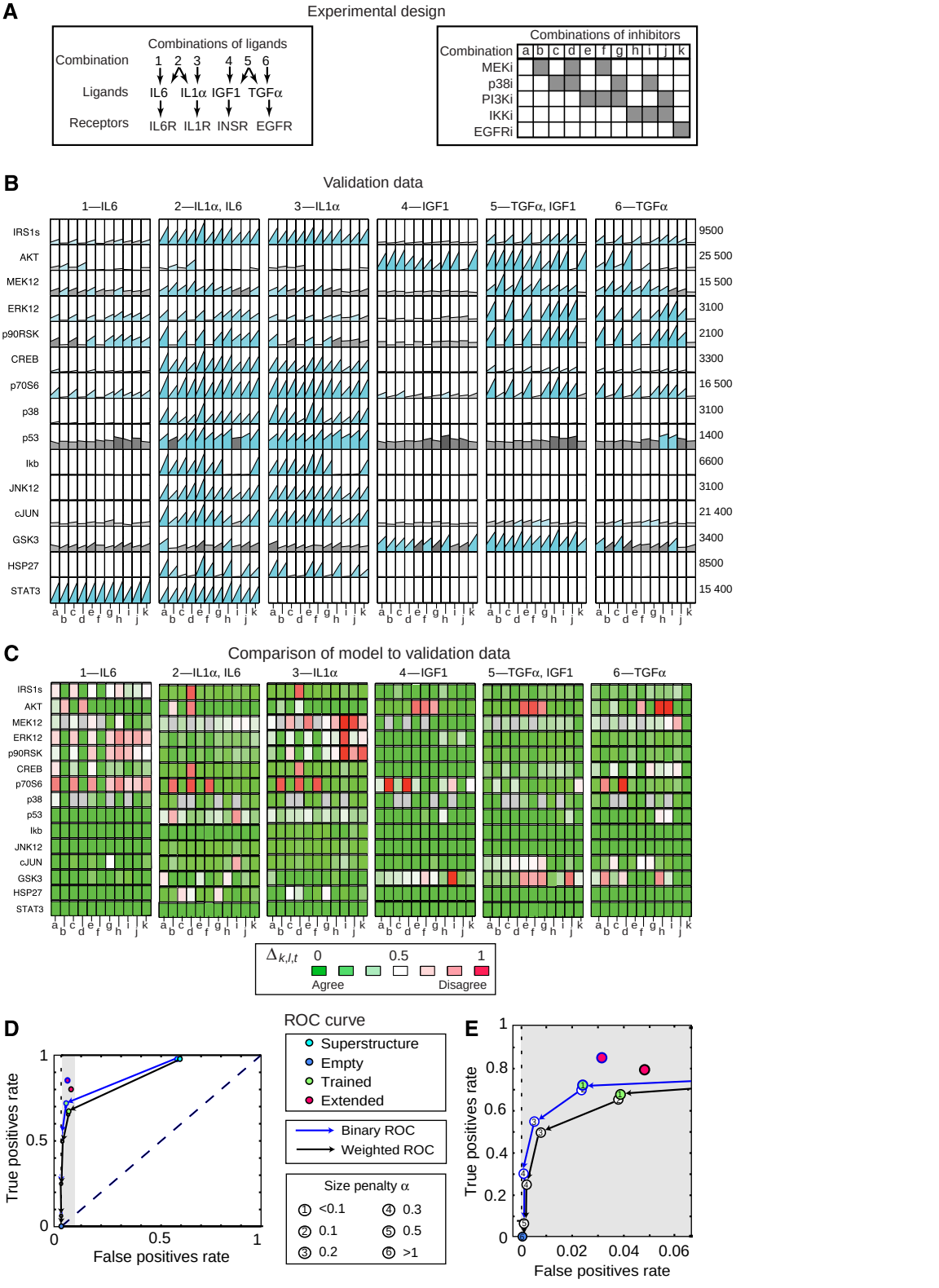
In this paper we describe a method to combine literature knowledge about mammalian signal transduction with functional data on the responses of cells to extracellular ligands and small-molecule drugs. Literature knowledge, in our case, comprised a signed directed graph assembled by manual curation of the literature (a PSN created using Ingenuity databases and software). In principle, however, any signed directed graph assembled from protein–protein interaction or gene association data could be used. To train network graphs against data we developed interoperable CNO and DataRail software that performs five essential tasks: (i) transforming graphs into compressed Boolean logic superstructures that can be used to compute input–output relationships for the overall network while containing the minimum number of non-identifiable elements; (ii) normalizing biochemical data on the states and activities of signalling proteins so that they can be used to train discrete two-state models; (iii) calibrating models to data based on an objective function that balances goodness of fit with model complexity; (iv) identifying new links not present in the starting graph that improve fit to data while marginally increasing model size and false-positive rate; and (v) manipulating calibrated models to enable their comparison to the starting graph. In addition to the CNO-based workflow for model assembly and calibration (Figure 2) we also describe a series of computational procedures, involving data and network randomization, derivation of Pareto frontiers, and computation of ROC curves that serve as tests of the quality and reliability of the modelling process.

We selected Boolean logic as the basis for the current work because it is the simplest form of logical modelling and involves no free parameters (for a given topology). Boolean logic has previously been applied to modelling cell signalling pathways, but the calibration of Boolean models to biological data has not been described. It was, therefore, unclear at the outset whether our approach would yield stable models with good predictive capability. Specifically, we were concerned that by assuming only two activity states for each element in the network (on and off), fitting errors would be too large to distinguish differences in pathway topology. However, we found that calibration of a complex cell signalling network, involving 82 species and 116 protein–protein interactions, against a relatively rich set of ~1000 protein measurements yielded a family of Boolean models with good fit to experimental data and relatively identifiable topologies. Evidently, the crudeness of the two-state representation is balanced by the feasibility of constructing a well-behaved objective function and performing many training runs.

We find that calibrated Boolean models contain fewer links than the PSNs on which they are based: whereas the starting graph for our network of seven receptors and 82 proteins had 1.42 interactions per node, trained Boolean models averaged 0.9 links per node. The explanation for the relatively poor fit to data a Boolean superstructure encoding the PSN is the large number of false-positive predictions so that even a completely empty model (one with no interactions) had a better overall fit. Nonetheless, the PSN used as a starting point for the current work is sparse compared with recently described networks. For example, a network of signalling pathways in neurons contained 545 nodes and 1259 interactions (2.3 edges per node; Ma'ayan *et al*, 2005), and undirected PINs and interactomes are even richer in edges per node: 3.5. for a network involving 121 disease-associated proteins (Rual *et al*, 2005), and 4.6 for a network constructed around tumour suppressors of breast cancer (Pujana *et al*, 2007). We therefore speculate that significant reductions in network complexity will be observed by applying the Boolean modelling approach described in this paper to PSNs and PINs covering other aspects of eukaryotic biology.

At least three explanations can be advanced to explain the reduction in the number of interactions per node, and the improvement in ROC characteristics, between starting PSNs and calibrated Boolean models: (i) interactions in the PSN are culled from several organisms, cell types, and growth

conditions, and only a subset of these interactions are relevant to a single cell type exposed to a limited set of ligands; (ii) interactions in the PSN are collapsed in time so that immediate-early and late events are indistinguishable, whereas our Boolean model is relevant only for events occurring within 30 min of ligand addition; and (iii) some interactions in PSNs are incorrect and—being based on two-hybrid, co-purification or other interaction assays—



overestimate the number of functional connections between proteins. In the future, it should be possible to distinguish among these possibilities by collecting data at multiple time points from multiple cell lines and then comparing a series of optimized models for each time point and cell type.

Comparing approaches to modelling complex protein networks

Reverse engineering has received considerable attention as a means to infer the topology of biological networks directly from patterns of co-variation in data. The use of prior knowledge distinguishes our work from standard reverse engineering (Sachs *et al*, 2005; Bonneau *et al*, 2006; Perkins *et al*, 2006; Bansal *et al*, 2007; Cho *et al*, 2007; Nelander *et al*, 2008), including reverse engineering using Boolean logic (D'haeseleer *et al*, 2000; Laubenbacher and Stigler, 2004). Moreover, we include in our decompressed graphs intermediate species that are not subject to experimental measurement, whereas latent or hidden variables are rarely included in reverse-engineered networks because they are supported only on the basis of prior knowledge. The absence of prior assumption in reverse-engineered networks is usually regarded as a significant advantage. However, our use of prior knowledge in the form of a PSN substantially improves the modelling process. This point is emphasized by the analysis of scrambled experimental data or randomized PSNs, neither of which yields models with as high a probability as real data and literature-based PSNs. Moreover, given the amount of biochemical and structural data available on mammalian signal transduction proteins, it seems unnecessarily restrictive to ignore prior knowledge completely. Doing so decouples reverse-engineered networks from mechanistic understanding and vastly increases the demands for rich multi-factorial data. Perhaps this explains why pure reverse engineering of mammalian networks has been most successful with networks containing <20 nodes.

At the other extreme, attempts to assemble accurate pictures of signal transduction based entirely on literature mining result in amalgamated maps—derived from many different cell types and even different organisms—that do not accurately describe or predict the behaviour of particular cells. Here we show that a middle road exists between the extremes of literature curation and reverse engineering that starts with construction of a network graph from the literature followed by pruning the graph through calibration to yield a model that is predictive for a specific biological situation. Reverse engineering methods should be highly complementary to the methods described here. Reverse engineering can uncover interactions in data that are not present in the literature and might, therefore, be missed. The fact that we identify two interactions as improving fit to data supports the assumption that other unidentified links undoubtedly exist in our PSN.

Several useful extensions of the CNO software described in this paper should be feasible. First, the approach can be extended to handle data collected at different time points. Simulation based on the computation of logical pseudo-steady state might not be the most appropriate way to capture the causality of dynamics of processes (e.g., it cannot describe phenomena such as oscillations), but it can be used to model reactions that operate on different time scales (Klamt *et al*, 2006). Second, we are implementing tools that extract PSNs from public sources such as Pathway Commons and to add data from PINs (Rual *et al*, 2005). These undirected graphs can serve as a source of information for adding edges to optimized models, but in this case we would need to evaluate four possible edges for each undirected edge, representing two directions and two signs. The different 'quality' of the prior knowledge can also be encoded by assigning different weights to the edges, as we do to search for links absent in the starting PSN.

A third area for development is optimizing the design of experiments that increase model identifiability. The Boolean model in the current work is non-identifiable given the available data, and this seems likely to be true of all similarly complex models of mammalian biology given practical constraints on experimentation. However, it will undoubtedly be possible to improve the modelling of certain nodes and edges by choosing the right combinations of biological stimuli, small-molecule inhibitors, and protein measurements. Further exploration of the landscape of the objective function, based for example on synthetic data derived from the model described here, should yield valuable insights into this issue.

Fourth, it should be possible to replace the deterministic approach used here with a more rigorous probabilistic approach in which each interaction is associated with a *P*-value. This *P*-value would derive both from the calibration process, in which case it would reflect the identifiability of the interaction based on the data, and also from the degree of confidence in the starting PSN. Bayesian approaches obviously represent an effective means to encode both the prior and consequent probabilities in such a network; progress in this direction can be found in the work of Gat-Viks and Shamir (2007). Moreover, probabilistic Boolean networks (Shmulevich *et al*, 2002) are a natural extension of the

Figure 7 Model validation involving prediction of a set of measurements not present in the training data. (**A**) Design of experiments in the validation data set depicting the use of four ligands in six combinations and five drugs in 11 combinations. Measurements were performed following the scheme outlined for the training data in Figure 4A. (**B**) The CSR data set obtained in HepG2 cells. Rows represent the measures of 15 intracellular signals assayed at the time of stimulation and 30 min later. For each combination of ligands and readout, a combination of the four inhibitors was used as described in panel A. Data are coded in blue to highlight induction. The data were processed using the DataRail software (Saez-Rodriguez *et al*, 2008). (**C**) Comparison of model prediction $B^{M}_{k,l,t}$ to normalized experimental data $B^{E}_{k,l,t}$. If the absolute value of the difference $\Delta_{k,l,t}=|B^{M}_{k,l,t}-B^{E}_{k,l,t}|=1$ (strongest disagreement), the corresponding box is coloured in red; if $\Delta_{k,l,t}=0$ (best agreement) it is in green; if $\Delta_{k,l,t}=0.5$ it is in white. Intermediate values of $\Delta_{k,l,t}$ were mapped to shades of red and green as shown. (**D**, **E**) The ROC curve of the trained model showing the ratio of true positives (1−the ratio of false negatives) and false positives. In panel E, the region of the ROC curve between 0 and 0.07 false-positive rate is shown in expanded form for clarity. The dots in the black curve correspond to the binary rates for models recovered over a range of size penalties (from $\alpha=0$ to 10, keyed to the legend). The complete superstructure is designated as a blue circle. The set of models shown in Figure 5 ($\alpha=0.0001$) is marked with a green circle and the extended model having two additional inferred links, with a red circle. The blue lines and circles correspond to the weighted ratio of false positives and negatives as described in the text. Source data is available for this figure at www.nature.com/msb.

deterministic Boolean networks used in this work. Finally, it may prove useful to add multilevel logic (Thomas and D'Ari, 1990), discrete Petri nets (Fisher and Henzinger, 2007; Chaouiya *et al*, 2008), Fuzzy-logic (Aldridge *et al*, 2009), or Boolean-based ODE systems (Mendoza and Xenarios, 2006; Wittmann *et al*, 2009) that can encode intermediate levels of protein activity. Some of these extensions will also make it possible to describe the dynamical process more accurately. Looking forward we anticipate that Boolean models of specific cell types should be useful in interpreting genetic data obtained from patients. PINs have been shown to improve the predictivity of gene expression classifiers used to discriminate disease states (Chuang *et al*, 2007), and it seems highly likely that logical models specific to a disease will prove even more effective in this role.

Materials and methods

CellNetOptimizer

CNO is a stand-alone Toolbox implemented in MATLAB that executes the Boolean logic and calibration procedures described in this paper. It can be used alone or in combination with DataRail, which manages experimental data according to a previously published MIDAS standard (Saez-Rodriguez *et al*, 2008). All functions are accessible either via scripting or graphical user interfaces. CNO can import models from ProMoT (Saez-Rodriguez *et al*, 2006) and CellNetAnalyzer (Klamt *et al*, 2007). In the near future, automatic population of CNO models from graphs in the BioPAX format (http://www.biopax.org/) and those stored in databases such as Pathway Commons, (http://www.pathwaycommons.org/) will be implemented. The models generated from CNO can be stored in DataRail as a data array, making it possible to store models alongside the data used for training. CNO is freely available at http://www.cdpcenter.org/resources/software/cellnetoptimizer/.

Model formalism

We use the Boolean modelling formalism as introduced by Klamt *et al* (2006) for modelling signal transduction networks. Nodes in the Boolean network represent biological species and have an associated logical value ('on' (1) or 'off' (0)) determining whether the species is active/present or not. The signalling events are encoded by Boolean operations on the network nodes. We describe the Boolean functions using the sum-of-products (SOP; also called the (minimal) disjunctive normal form (DNF)) representation (Mendelson, 1970) that uses only AND, OR, and NOT operators. A SOP expression is a sum (i.e., OR connection) of terms where each term is either a single, possibly negated Boolean variable or a product (i.e., AND connection) of (possibly negated) Boolean variables. Any logical operation can be represented in this way. For example, an XOR gate is described in the SOP formalism as (A is ON and B is OFF) OR (A is OFF and B is ON).

A Boolean network in which Boolean functions are given as SOPs can be represented as a directed hypergraph. A directed hypergraph consists of a set of nodes and a set of directed hyperedges. Each hyperedge connects two sets of nodes, the tail (containing the start nodes) and the head (containing the end nodes). Tail and head can contain several nodes, although in this paper they have only one head (end) node. A conventional graph is simply a special case of a hypergraph in which the cardinality of the tails and heads is 1 for all edges (Klamt *et al*, 2009). By using the SOP formalism, a logic network can be converted to a hypergraph in a straightforward manner. Each hyperedge pointing into node i represents one term of the Boolean function (i.e., an AND connection or a single Boolean variable) describing the activation mechanism of species i and thus represents one way of activating the node. All hyperedges ending in a node are implicitly linked via OR logic (Klamt *et al*, 2006).

Network preprocessing

An implementation of the Floyd–Warshall algorithm (Floyd, 1962), drawn from CellNetAnalyzer, is used to find paths among species. This algorithm makes it possible to identify non-observable and non-controllable elements: if no path can be found from a species (node) to any readout, the species is non-controllable; if no path can be found from any cue (stimulus or inhibitor) to a species, the species in non-observable.

Model simulation and comparison to experimental data

Based on time-resolved experiments, we identified 30 min after ligand stimulation as the time point at which phosphorylation levels differed maximally from those of untreated controls (time point 0). We assumed these values reflected a state achieved on a time scale on which fast events are relevant, but slow events (such as protein degradation) have a relatively insignificant effect. Qualitatively, these states can be computed as logical steady states in the Boolean network describing the early events (Klamt *et al*, 2006). We therefore computed, for each model candidate, the logical steady state associated with the input values determined by each experiment k and compared the values of the readouts $B^{M}_{k,l,30}$ with the normalized experimental values $B^{E}_{k,l,30}$ using the MSE deviation as explained in the main text.

We compute the logical steady state resulting from the input stimuli by propagating input signals along logical (hyperarc) connections (Klamt *et al*, 2006). Whether or not we can resolve a complete and unique logical response of all nodes for a given set of input stimuli, depends on the functionality of positive- or negative-feedback loops in the network (e.g., negative-feedback loops may prevent the establishment of a logical steady state). If the state of a readout is undetermined (i.e., if no unique logical response for this node can be resolved), the resulting model is penalized as if it were incorrectly predicting the data for that experiment. The simulation can be extended to multiple time points by considering that each time point is a characteristic time scale where a certain pseudo-steady state is reached; however, this is not implemented in the current work. Boolean models can be used to analyse cyclic attractors, using either synchronous or asynchronous updates (Thomas and D'Ari, 1990). Cyclic attractors are associated with oscillatory behaviour, which is absent from our data set, and we have, therefore, not explored the use of CNO with cyclic attractors yet. Each node has an associated default value corresponding to the 'resting' network (no stimuli present), which is 0 (inactive) for all nodes except for IκB and GSK3 (which act as negative regulators that are on at the start of the experiment). The value of a node that has no input is given by its default value, but the value of all other nodes is overwritten by signals propagated from the inputs.

Computer routines to perform simulations (in particular, to compute logical steady states that are generated by a certain combination of stimuli and inhibitors) are original to this work or are adapted from CellNetAnalyzer (Klamt *et al*, 2007).

Reduction of search space using Sperner systems

For n nodes upstream of a given node (with fixed sign for each edge), there are $h(n)=2^n-1$ possible hyperedges (one for every subset of $\{1,\ldots,n\}$ nodes minus the empty set). Combining h hyperedges, we can construct $g(n) = 2^h = 2^{2^n-1}$ Boolean functions; each of these functions can be represented by a binary vector indicating which hyperedge is part of the function (1) and which is not (0). $g(n)$ corresponds to the number of possible vectors of length h. However, many of those Boolean functions will be redundant in the sense that some of them have the same truth table.

As an example, let nodes X and Y lie upstream of node A, that is, X and Y are predecessors of A. Then we have $n=2$ inputs from which we can construct $h=3$ hyperedges: ($h1$) X→A, ($h2$) Y→A, and ($h3$) (X AND Y)→A. One function we can construct consists of ($h1$) X→A plus ($h3$) (X AND Y)→A (that is,'X OR (X AND Y) leads to A'). However, this Boolean function has the same truth table as the one consisting solely of ($h1$) X→A. Overall, only five possible truth tables exist despite the presence of eight Boolean functions, making three of them redundant

($h1+h2$, $h1+h3$, and $h1+h2+h3$). Optimization of the objective function (equation (1) in the main text) is made considerably more efficient by omitting redundant Boolean functions.

Formally, in the SOP representation, redundant Boolean functions may occur if the terms (the AND connections corresponding to the hyperedges) contain more variables than necessary and can thus be simplified by removing some of the variables contained in them. Irreducible terms are called prime implicants. In the example, 'X OR (X AND Y) leads to A' is redundant as it can be replaced by 'X leads to A' (the term 'X AND Y' is not a prime implicant because 'X' is a subset). Accordingly, we say that a set of hyperedges is non-redundant if it encodes a non-redundant Boolean function (consisting only of prime implicants) and this is true if and only if there is no hyperedge whose tail contains a tail of another hyperedge, that is, if they form a Sperner system (Bollobas, 1986). Therefore, during the optimization routine, instead of checking all subsets of possible hyperedges ($2^h = 2^{2^n-1}$), we restrict ourselves to checking only those that form a Sperner system. There is no general expression for the number $S(n)$ of Sperner systems, but as an example, for n=1,2,3,4,5, the number of Sperner systems $S(n)$ are $S(1)$=1, $S(2)$=4, $S(3)$=18, $S(4)$=166, and $S(5)$=7579. In contrast, the number of all SOP representations of the Boolean functions is $g(1)$=2, $g(2)$=8, $g(3)$=128, $g(4) \approx 3 \cdot 10^4$, and $g(5) \approx 2 \cdot 10^9$.

We have developed an extension of our optimization procedure that considers Sperner systems within CNO. To implement this concept, CNO defines a vector S. Each element s_i can have a value 0, 1,…, $S(n_i)$, where $S(n_i)$ is the number of Sperner systems for the node i. Each value of the vector S can be mapped to a value of the vector P in equation (1), so that the Sperner hypergraphs can be evaluated and therefore the optimal model can be identified.

In the networks in this paper, the use of Sperner systems reduced the search space to approximately the square root of the original size.

Genetic algorithm for optimization

To search over the possible models encoded in the superstructure when enumeration is not feasible, we implemented in CNO a previously described genetic algorithm (Goldberg, 1989) according to the following rules:

(1) **Start:** A population of model variants (encoded in vectors P; see equation (1)) is initialized. We explored different initialization strategies, including random networks, a full superstructure, and an empty model, obtaining the same results for all cases.
(2) **Fitness**: The fitness of each individual (model variant encoded in a vector P) is determined as a function of its objective function Θ. We explored two methods to assign fitness: Ranking, where the fitness is based on the rank of the individual in the population in terms of Θ and Proportional with sigma scaling, where the value is proportional to Θ and scaled to the standard deviation to avoid premature convergence. Preliminary studies showed both methods to yield similar results, but Ranking was slightly better, and that is what we chose.
(3) **Generation of new population**: We used the following steps: (a) Selection: Individuals are selected from the population to reproduce, assigning a higher chance of reproduction to individuals with higher fitness using Stochastic Uniform Sampling (Mitchell, 1998). (b) Crossover: Individuals mate (following uniform crossover) so that the offspring inherit certain parts of the vector P from each parent. (c). Mutation: Individuals can mutate at specific loci in the chromosome (vector P). We explored different values for the mutation probability without an effect on the solution; most results were obtained with a probability 0.5.
(4) **Replace**: A new population replaces the old one. We implemented elitism, where the five best individuals of each generation were directly passed onto the next generation.
(5) **Test stop criteria**: Several stop criteria are checked for each generation, including tolerance from a perfect fit, as well as number of generations without improvements in the fit of the best individual (stall generations). We chose a number of stall generations (10 000) large enough to make sure that the solution reached was stable.
(6) **Loop**: If any of the stop criteria are fulfilled, the optimization stops. Otherwise, it iterates to step 2.
(7) **Post-processing:** A genetic search does not necessarily yield the lowest value of the objective function, so a post-processing step is performed where individual interactions are pruned: We evaluate exhaustively the effect that removal of individual interactions has on the value of the objective function. If the fit to data does not get worse, the interaction is removed from the final solution to minimize model size.

Construction of a signed directed graph of growth and inflammatory signalling

Our model of inflammatory and growth signalling pathways in liver started with a graph assembled from pathways in the Ingenuity IPA software (www.ingenuity.com) that summarize the relevant literature (see Supplementary Table 1 for the full names of the abbreviations). As it is based largely on biochemical and molecular data, the Ingenuity-derived graph is signed and directed. However, the description of IRS-1 biology in Ingenuity is poor and we, therefore, added additional information from the literature as follows: the species IRS-1 as described in Ingenuity was considered to describe tyrosine phosphorylation of IRS-1, and was renamed accordingly as IRS-1Y. The activation of IRS-1Y, which is dependent on IGF-1 stimulation, was considered to be inhibited if the serine site was phosphorylated (Hotamisligil *et al*, 1996; Saltiel and Kahn, 2001). We therefore added a node (IRS-1S) for the serine site of IRS-1. IRS-1S, in turn, was considered to be dependent on mTOR activation (Ozes *et al*, 2001).

Data generation

The design and execution of multiplex experiments is described elsewhere (Alexopoulos *et al*, 2009, in preparation). Briefly, HepG2 cells were plated in 96-well plates coated with collagen type-I (Becton Dickinson), with 100 μl phenol-free Williams' Medium E (WEM; Sigma-Aldrich) with media supplements and fetal calf serum. Cells were cultured overnight on collagen, starved for 6 h in 180 μl of WEM lacking serum, and then exposed to kinase inhibitors and ligand cues. Cells were lysed in 90 μl of manufacturer's lysis buffer (Bio-Rad) and intracellular signals were measured using high-throughput sandwich immunoassays (Luminex xMAP assay; Austin, TX, USA). Specifically, a 17-plex phospho-protein bead set from Bio-Rad was used to assay the phosphorylation of the following proteins: p70S6K (T421/S424), CREB (S133), p90RSK (T359/S363), p38 (T180/Y182), MEK1 (S217/S221), JNK (T183/Y185), HSP27 (S78), ERK1/2 (T202/Y204, T185/Y187), c-Jun (S63), IRS-1 (S636/S639), STAT3 (Y705), IκB-α (S32/S36), histone H3 (S10), p53 (S15), GSK-3α/β (S21/S9), and AKT (S473). The training data set can be downloaded as a MIDAS file (Saez-Rodriguez *et al*, 2008) from http://www.cdpcenter.org/resources/data/alexopoulos-et-al-2009/ and the test data set from http://www.cdpcenter.org/resources/data/saez-rodriguez-et-al-2009/ or from the article's webpage.

Reagents

Ligand cues

TNFα, IGF-1, and TGFβ1 were obtained from PeproTech; LPS and IL-6 were from Sigma-Aldrich; IL-1α and TGFα were from R&D Systems; and IFN-γ (hIFN-γ) was from Roche Diagnostics GmbH. Other than LPS, TLR ligands were obtained from InvivoGen as follows: Pam3CSK4 for TLR1/2; HKLM for TLR2; poly(I:C) for TLR3; *Salmonella typhimurium* flagellin for TLR5; FSL1-Pam2CGDPKHPKSF for TLR6/2; imiquimod for TLR7; ssRNA40 for TLR8; and ODN2006 for TLR9.

Kinase inhibitors

Inhibitors for IKK2 (BMS-345541), PI3K (ZSTK474), GSK3β (inhibitor XI), JNK (SP600125), and mTOR (rapamycin) were purchased from Calbiochem.

Inhibitors for p38 (PHA818637) and MEK (PD325901) were kindly provided by Pfizer Pharmaceuticals.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank WW Chen, M Niepel, M Morris, J Wagner, S Hautaniemi, K Jaqaman, N Domedel-Puig, F Theis, J Bernanke, P Vera-Licona, E Sontag, and ED Gilles for useful discussions; J Muhlich, L Kleiman, A Goldsipe, and S Mirschel for scientific and technical assistance; and members of the Pfizer Research Technology Center for support in the early phases of the project. This work was funded by NIH grants P50-GM68762 and U54-CA112967 to PKS and DAL. RS and SK are funded by the German Federal Ministry of Education and Research ('HepatoSys' and the FORSYS-Centre MaCS).

Conflict of interest

The authors declare that they have no conflict of interest.

References

Akaike H (1974) A new look at the statistical model identification. *Automat Contr* **19:** 716–723

Aldridge B, Saez-Rodriguez J, Muhlich J, Sorger P, Lauffenburger DA (2009) Fuzzy logic analysis of kinase pathway crosstalk inTNF/EGF/insulin-induced signaling. *PLoS Comput Biol* **5:** e1000340

Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK (2006) Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* **8:** 1195–1203

Bader G, Cary MP, Sander C (2006) Pathguide: a pathway resource list. *Nucleic Acids Res* **34:** D504–D506

Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Mol Syst Biol* **3:** 78

Barabási AL, Oltvai Z (2004) Network biology: understanding the cell's functional organization. *Nat Rev Genet* **5:** 101–113

Barron A, Rissanen J, Yu B (1998) The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory* **44:** 2743–2760

Bauer-Mehren A, Furlong LI, Sanz F (2009) Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol* **5:** 290

Bollobas B (1986) *Combinatorics: Set Systems, Hypergraphs, Families of Vectors, and Combinatorial Probability*. Cambridge, UK: Cambridge University Press

Bonneau R, Reiss DJ, Shannon P, Facciotti M, Hood L, Baliga NS, Thorsson V (2006) The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets *de novo*. *Genome Biol* **7:** R36

Bossi A, Lehner B (2009) Tissue specificity and the human protein interaction network. *Mol Syst Biol* **5:** 260

Chaouiya C, Remy E, Thieffry D (2008) Petri net modelling of biological regulatory networks. *J Disc Algorithms* **6:** 165–177

Chatr-Aryamontri A, Ceol A, Palazzi L, Nardelli G, Schneider M, Castagnoli L, Cesareni G (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res* **35:** D572–D574

Chaves M, Albert R, Sontag ED (2005) Robustness and fragility of Boolean models for genetic regulatory networks. *J Theoret Biol* **235:** 431–449

Cho KH, Choo SM, Jung SH, Kim JR, Choi HS, Kim J (2007) Reverse engineering of gene regulatory networks. *IET Syst Biol* **1:** 149–163

Chuang H, Lee E, Liu Y, Lee D, Ideker T (2007) Network-based classification of breast cancer metastasis. *Mol Syst Biol* **3:** 10

Cusick M, Yu H, Smolyar A, Venkatesan K, Carvunis A, Simonis N, Rual J, Borick H, Braun P, Dreze M, Vandenhaute J, Galli M, Yazaki J, Hill D, Ecker J, Roth FP, Vidal M (2009) Literature-curated protein interaction datasets. *Nat Methods* **6:** 39–46

D'haeseleer P, Liang S, Somogyi R (2000) Genetic network inference: from co-expression clustering to reverse engineering. *Bioinformatics* **16:** 707–726

de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* **9:** 67–103

Fanger GR, Johnson NL, Johnson GL (1997) MEK kinases are regulated by EGF and selectively interact with Rac/Cdc42. *EMBO J* **16:** 4961–4972

Fauré A, Naldi A, Chaouiya C, Thieffry D (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* **22:** e124–e131

Fisher J, Henzinger T (2007) Executable cell biology. *Nate Biotechnol* **25:** 1239–1249

Floyd RW (1962) Algorithm 97 (SHORTEST PATH). *Commun ACM* **5:** 345

Gat-Viks I, Shamir R (2007) Refinement and expansion of signaling pathways: the osmotic response network in yeast. *Genome Res* **17:** 358–367

Gaudet S, Janes KA, Albeck JG, Pace EA, Lauffenburger DA, Sorger PK (2005) A compendium of signals and responses triggered by prodeath and prosurvival cytokines. *Mol Cell Proteomics* **4:** 1569–1590

Goldberg DE (1989) *Genetic Algorithms in Search, Optimization, and Machine Learning*. Boston, MA, USA: Addison-Wesley Professional

Guan KL, Figueroa C, Brtva TR, Zhu T, Taylor J, Barber TD, Vojtek AB (2000) Negative regulation of the serine/threonine kinase B-Raf by Akt. *J Biol Chem* **275:** 27354–27359

Gupta S, Bisht SS, Kukreti R, Jain S, Brahmachari SK (2007) Boolean network analysis of a neurotransmitter signaling pathway. *J Theoret Biol* **244:** 463–469

Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* **100:** 57–70

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, Eilbeck K, Lewis S, Marshall B, Mungall C, Richter J, Rubin GM, Blake JA, Bult C, Dolan M, Drabkin H, Eppig JT, Hill DP, Ni L, Ringwald M *et al* (2004) The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res* **32:** D258–D261

Herbst RS (2002) ZD1839: targeting the epidermal growth factor receptor in cancer therapy. *Expert Opin Invest Drugs* **11:** 837–849

Hers I, Tavaré JM (2005) Mechanism of feedback regulation of insulin receptor substrate-1 phosphorylation in primary adipocytes. *Biochem J* **388:** 713–720

Hotamisligil GS, Peraldi P, Budavari A, Ellis R, White MF, Spiegelman BM (1996) IRS-1-mediated inhibition of insulin receptor tyrosine kinase activity in TNF-alpha- and obesity-induced insulin resistance. *Science* **271:** 665–668

Hotte SJ, Hirte HW (2002) BAY 43-9006: early clinical data in patients with advanced solid malignancies. *Curr Pharm Des* **8:** 2249–2253

Huang S, Ingber DE (2000) Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks. *Exp Cell Res* **261:** 91–103

Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M (2003) A Bayesian networks approach for predicting protein–protein interactions from genomic data. *Science* **302:** 449–453

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M (2004) The KEGG resource for deciphering the genome. *Nucleic Acids Res* **32:** D277–D280

Kauffman SA (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theoret Biol* **22:** 437–467

Kerrien S, Alam-Faruque Y, Aranda B, Bancarz I, Bridge A, Derow C, Dimmer E, Feuermann M, Friedrichsen A, Huntley R, Kohler C, Khadake J, Leroy C, Liban A, Lieftink C, Montecchi-Palazzi L, Orchard S, Risse J, Robbe K, Roechert B *et al* (2007) IntAct—open source resource for molecular interaction data. *Nucleic Acids Res* **35:** D561–D565

Klamt S, Haus UU, Theis F (2009) Hypergraphs and cellular networks. *PLoS Comput Biol* **5:** e1000385

Klamt S, Saez-Rodriguez J, Gilles ED (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol* **1:** 2

Klamt S, Saez-Rodriguez J, Lindquist J, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* **7:** 56

Köcher T, Superti-Furga G (2007) Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nat Methods* **4:** 807–815

Kremling A, Saez-Rodriguez J (2007) Systems biology—an engineering perspective. *J Biotechnol* **129:** 329–351

Laubenbacher R, Stigler B (2004) A computational algebra approach to the reverse engineering of gene regulatory networks. *J Theoret Biol* **229:** 523–537

Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431:** 308–312

Ma'ayan A, Jenkins S, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, Kershenbaum A, Stolovitzky GA, Blitzer RD, Iyengar R (2005) Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science* **309:** 1078–1083

MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. *Science* **289:** 1760–1763

Marchetti L, Klein M, Schlett K, Pfizenmaier K, Eisel UL (2004) Tumor necrosis factor (TNF)-mediated neuroprotection against glutamate-induced excitotoxicity is enhanced by *N*-methyl-D-aspartate receptor activation. Essential role of a TNF receptor 2-mediated phosphatidylinositol 3-kinase-dependent NF-kappa B pathway. *J Biol Chem* **279:** 32869–32881

Mendelson E (1970) *Schaum's Outline of Boolean Algebra and Switching Circuits*. New York: McGraw-Hill

Mendoza L, Xenarios I (2006) A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor Biol Med Modelling* **3:** 13

Mitchell M (1998) *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press

Nelander S, Wang W, Nilsson B, She Q, Pratilas C, Rosen N, Gennemark P, Sander C (2008) Models from experiments: combinatorial drug perturbations of cancer cells. *Mol Syst Biol* **4:** 11

Ozes ON, Akca H, Mayo LD, Gustin JA, Maehama T, Dixon JE, Donner DB (2001) A phosphatidylinositol 3-kinase/Akt/mTOR pathway mediates and PTEN antagonizes tumor necrosis factor inhibition of insulin signaling through insulin receptor substrate-1. *Proc Natl Acad Sci USA* **98:** 4640–4645

Perkins TJ, Jaeger J, Reinitz J, Glass L (2006) Reverse engineering the gap gene network of *Drosophila melanogaster*. *PLoS Comput Biol* **2:** e51

Pieroni E, de la Fuente van Bentem S, Mancosu G, Capobianco E, Hirt H, de la Fuente A (2008) Protein networking: insights into global functional organization of proteomes. *Proteomics* **8:** 799–816

Pujana M, Han J, Starita L, Stevens K, Tewari M, Ahn J, Rennert G, Moreno V, Kirchhoff T, Gold B, Assmann V, Elshamy W, Rual J, Levine D, Rozek L, Gelman R, Gunsalus K, Greenberg R, Sobhian B, Bertin N *et al* (2007) Network modeling links breast cancer susceptibility and centrosome dysfunction. *Nat Genet* **39:** 1338–1349

Rhee SH, Keates AC, Moyer MP, Pothoulakis C (2004) MEK is a key modulator for TLR5-induced interleukin-8 and MIP3alpha gene expression in non-transformed human colonic epithelial cells. *J Biol Chem* **279:** 25179–25188

Rual J, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N, Klitgord N, Simon C, Boxem M, Milstein S, Rosenberg J, Goldberg DS, Zhang LV, Wong SL, Franklin G, Li S *et al* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature* **437:** 1173–1178

Russell M, Lange-Carter CA, Johnson GL (1995) Direct interaction between Ras and the kinase domain of mitogen-activated protein kinase kinase kinase (MEKK1). *J Biol Chem* **270:** 11757–11760

Sachs K, Perez O, Pe'er D, Lauffenburger DA, Nolan GP (2005) Causal protein-signaling networks derived from multiparameter single-cell data. *Science* **308:** 523–529

Saez-Rodriguez J, Goldsipe A, Muhlich J, Alexopoulos LG, Millard B, Lauffenburger DA, Sorger PK (2008) Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics* **24:** 840–847

Saez-Rodriguez J, Mirschel S, Hemenway R, Klamt S, Gilles ED, Ginkel M (2006) Visual setup of logical models of signaling and regulatory networks with ProMoT. *BMC Bioinformatics* **7:** 506

Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, Arndt B, Haus UU, Weismantel R, Gilles ED, Klamt S, Schraven B (2007) A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol* **3:** e163

Saltiel A, Kahn C (2001) Insulin signalling and the regulation of glucose and lipid metabolism. *Nature* **414:** 799–806

Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S (2009) The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comp Biol* **5:** e1000438

Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* **6:** 461–464

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498–2504

Shmulevich I, Dougherty ER, Kim S, Zhang W (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18:** 261–274

Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL (2009) Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol* **27:** 199–204

Thomas R, D'Ari R (1990) *Biological Feedback*. Boca Raton, FL, USA: CRC Press

Thomas R, Kaufman M (2001) Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos (Woodbury, NY)* **11:** 180–195

Tibshirani R (1994) Regression shrinkage and selection via the lasso. *J R Stat Soc B* **58:** 267–288

Wittmann D, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, Theis F (2009) From qualitative to quantitative modeling. *BMC Systems Biol* **3:** 98

Yaguchi S, Fukui Y, Koshimizu I, Yoshimi H, Matsuno T, Gouda H, Hirono S, Yamazaki K, Yamori T (2006) Antitumor activity of ZSTK474, a new phosphatidylinositol 3-kinase inhibitor. *J Natl Cancer Inst* **98:** 545–556

Zhang R, Shah M, Yang J, Nyland S, Liu X, Yun J, Albert R, Loughran T (2008) Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci USA* **105:** 16308

Zhao W, Serpedin E, Dougherty ER (2006) Inferring gene regulatory networks from time series data using the minimum description length principle. *Bioinformatics* **22:** 2129–2135

Zhou D, He Y (2008) Extracting interactions between proteins from the literature. *J Biomed Informatics* **41:** 393–407